MAYO CLINIC

Terminology Service Development at the Mayo Clinic

Harold Solbrig
Technical Specialist
Medical Informatics Research
Mayo Clinic
Rochester, MN 55901

MAYO CLINIC

Outline

- **Background** – why terminology services?
- **Where we've been** – the evolution of terminology services at the Mayo Clinic
- **Where we are now** – current tools and approaches to terminology services
- **Where we are going** – technologies, distribution, future applications

Background

MAYO CLINIC

Terminology

The lexicon of a special subject language reflects the organisational characteristics of the discipline by tending to provide as many lexical units as there are concepts…

Juan C. Sager, *A Practical Course in Terminology Processing.* John Benjamins, 1990

MAYO CLINIC

Terminology

The items which are characterised by special reference within a discipline are the 'terms' of the discipline, and collectively they form its 'terminology'; those which function in general reference over a variety of sublanguages are simply called 'words', and their totality the 'vocabulary'.

Juan C. Sager. *A Practical Course in Terminology Processing*

MAYO CLINIC

Evolutionary Steps of a Terminology

1) **Everyday words (vocabulary)** - the differentiating knowledge of the trade or profession is in the formative stage.

2) **Overloaded words and/or acronyms** - As the knowledge increases, it becomes cumbersome to continue to use full phrases to express concepts. Common phrases are shortened: "*kernel, heap, garbage collection, SCRAM, LASER, ...*"

MAYO CLINIC

Evolutionary Steps of a Terminology

3) **Formalized nomenclature** - as the need for precision and detail increases, the management of haphazard wording becomes prohibited. When practical, formalized naming rules are established. *Linnaeus system, chemical names, SNOMED, etc.*

4) **Coding and classification schemes** - as specialties and different views emerge, the need to classify, categorize and cross-reference becomes important. *ICD, etc.*

MAYO CLINIC

Evolutionary Steps of a Terminology

5) **Thesauri** - Boundaries between specialties change and cross references become necessary between terminologies of different specialties. *UMLS*

6) **Reference Terminologies** - Thesauri become unwieldy and too imprecise. A new, synthetic, atomic conceptual organization is formed as a reference point and focus. *SNOMED-CT, Read Codes III*

MAYO CLINIC

Terminological Categories

"First Generation" terminological systems

- Typically targeted for paper based information systems (not "IT-enabled")
- Simple hierarchies and organization
- Expensive to maintain and reuse

Angelo Rossi Mori, et. al. *Standards to support development of terminological systems for healthcare telematics.*

MAYO CLINIC

Terminological Categories

“Second Generation” systems

- Compositional systems, based on a *categorial structure* (semantic categories, semantic links [associations between categories])
- Dynamic organization, systematic description of a subject field
- Flexible and extensible
- Limited semantic based processing

MAYO CLINIC

Terminological Categories

- **"Third Generation" systems**
  - Based on a formal model (a set of symbols and a set of formal rules)
  - Dynamic w/ multiple hierarchies
  - Content and updates are formally validated
  - Complete semantic based processing (behavior is independent of names)

MAYO CLINIC

What is a terminology?

Key characteristics:

- Set of terms, definitions and relationships for a (relatively) non-ambiguous partitioning of the conceptual space of a specialized subject area or discipline.
- NOT necessarily related to computerized data processing (or even data processing, period)
- A formal shared context for communication among members of a specialty or trade.

What is a terminology?

- **Key characteristics**
  - Term ⟷ concept mapping
  - Additional entry phrases including
    - Lexical variants
    - Synonyms
    - Similar or related phrases
  - Intrinsic definitions, annotations, etc.
  - Extrinsic definitions in form of taxonomy / ontology / semantic net

MAYO CLINIC

Terminology vs. Ontology

- The word “Ontology”, as it is used today refers to the DL-based organization of ‘concepts’
- Focus is on the formal organization
- Lexical/linguistic section is underspecified
  - Attributed definitions
  - Terms in multi-languages and contexts
- Behavioral characteristics are strictly DL – no rules on how to find a node given an input string...

MAYO CLINIC

Uses of Terminology in Clinical Practice

MAYO CLINIC

Terminology in Clinical Practice

Code Sets – Lists of codes used to fill out forms, data entry, etc.

- Drawn from small to medium size lists
- Typically local to institution
- May not cross databases or applications

MAYO CLINIC

Code Sets in Forms

10. Complete the following information if the isolate is *vibrio cholerae 01 or 0139:*

Serotype (452) (check one)

_ Inaba (1)
_ Ogawa (2)
_ Hikojima (3)
_ Not Done (4)
_ Unk. (9)

Biotype (check one)

_ El Tor (1)
_ Classical (2)
_ Not Done (3)
_ Unk. (9)

Patient home state: _ _

4. Sex: (68)

_ M (1)
_ F (2)
_ Unk. (9)

The patient has been enrolled at:

1	NIH-sponsored
2	Other
3	None
9	Unknown

MAYO CLINIC

Terminology in Clinical Practice

Classifications - Codes used to summarize information for the purpose of QA, reporting, reimbursement, etc.

- ICD-9-CM, CPT4, ...
- (Usually) Redundant information
- In use since (at least) the 16th century London Bills of Mortality

Terminology in Clinical Practice

Classifications

- **Shape and size of "buckets" depends on intended use**
  - "Killed by several Accidents"
  - "King's Evil"
  - "Frightened"
  - "Crushed by falling aircraft in terrorism [attack]"
  - "concussion with more than 24 hours loss of consciousness and return to pre-existing conscious level"

MAYO CLINIC

Terminology in Clinical Practice

Indices – Codes used to summarize the content of medical records for the purpose of research and retrieval

- MESH, ICD-9-CM, HICDA, ...

MAYO CLINIC

Terminology in Clinical Practice

Metadata – Codes that describe the format and content of databases, files, forms, etc.

- Enables sharing of information across institutions
- Enables sharing of information within institutions across time
- Only recently becoming formalized

MAYO CLINIC

Terminology in Clinical Practice

- **Code Sets**
  - Often restrictive, incomplete
  - Lack of compositional structure
  - Not applicable in many settings (free text, quantitative data, etc.
- **Classifications**
  - Granularity depends on the context
  - Rarely matches the level of specificity needed to accurately record clinical information

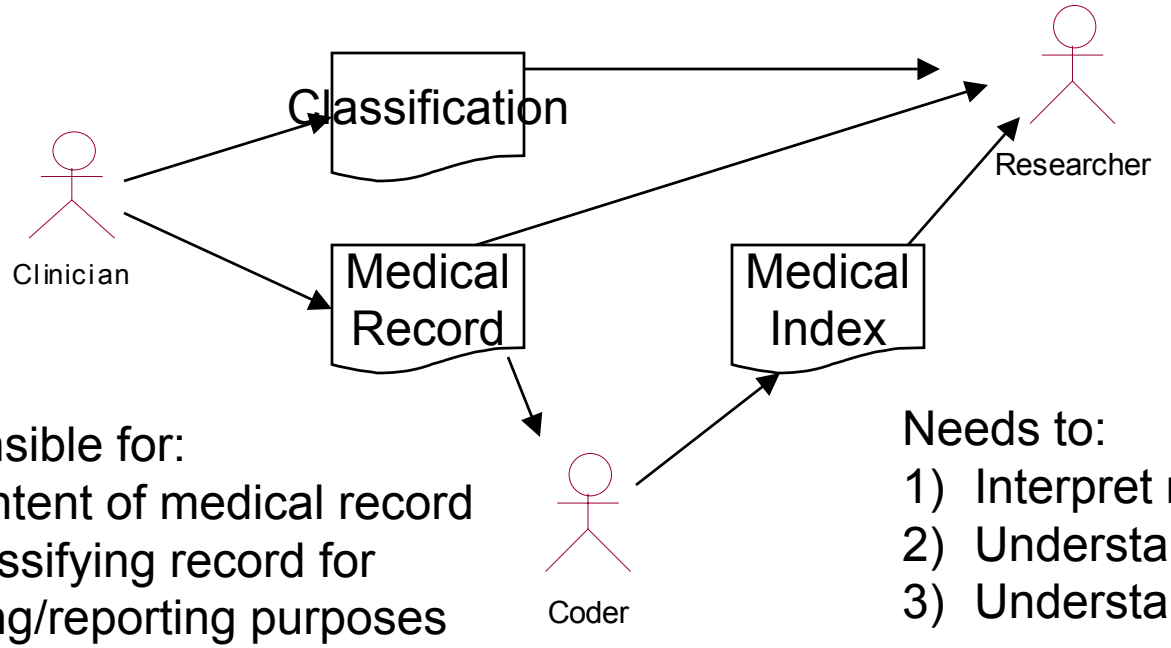
MAYO CLINIC

Terminology in Clinical Practice

- **Indices**
  - Classification after the fact
  - Can be labor intensive
  - Balance must be maintained between granularity and cost
  - Cannot anticipate unexpected requests
    - AIDS symptoms
    - Terrorism related events

MAYO CLINIC

Terminology in Clinical Practice



Responsible for:
1) Content of medical record
2) Classifying record for billing/reporting purposes

Needs to:
1) Interpret record contents
2) Understand classification process
3) Understand indexing process

Responsible for:
1) Interpreting medical record
2) Classifying record for research purposes

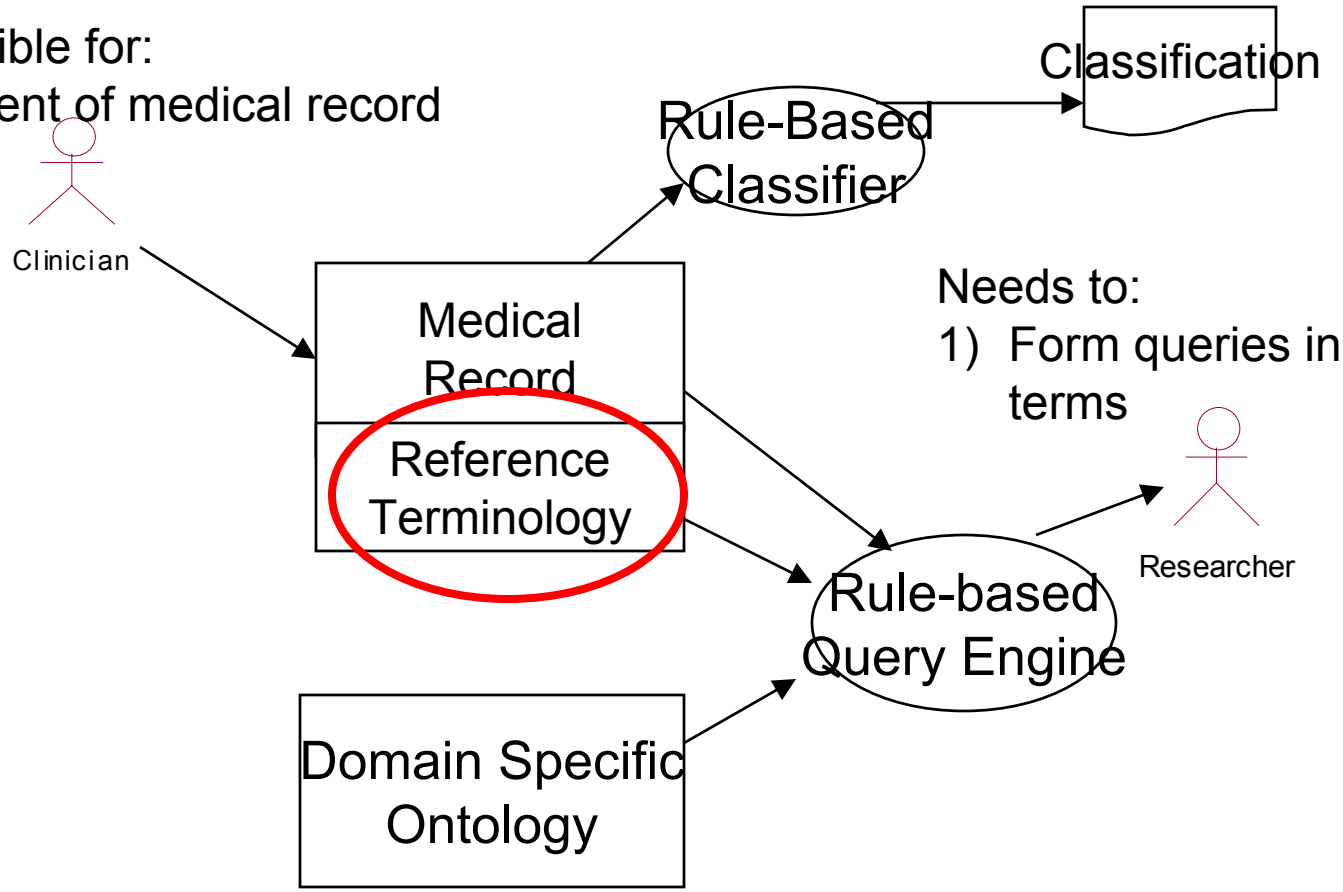
MAYO CLINIC

Some ways that the process can fail

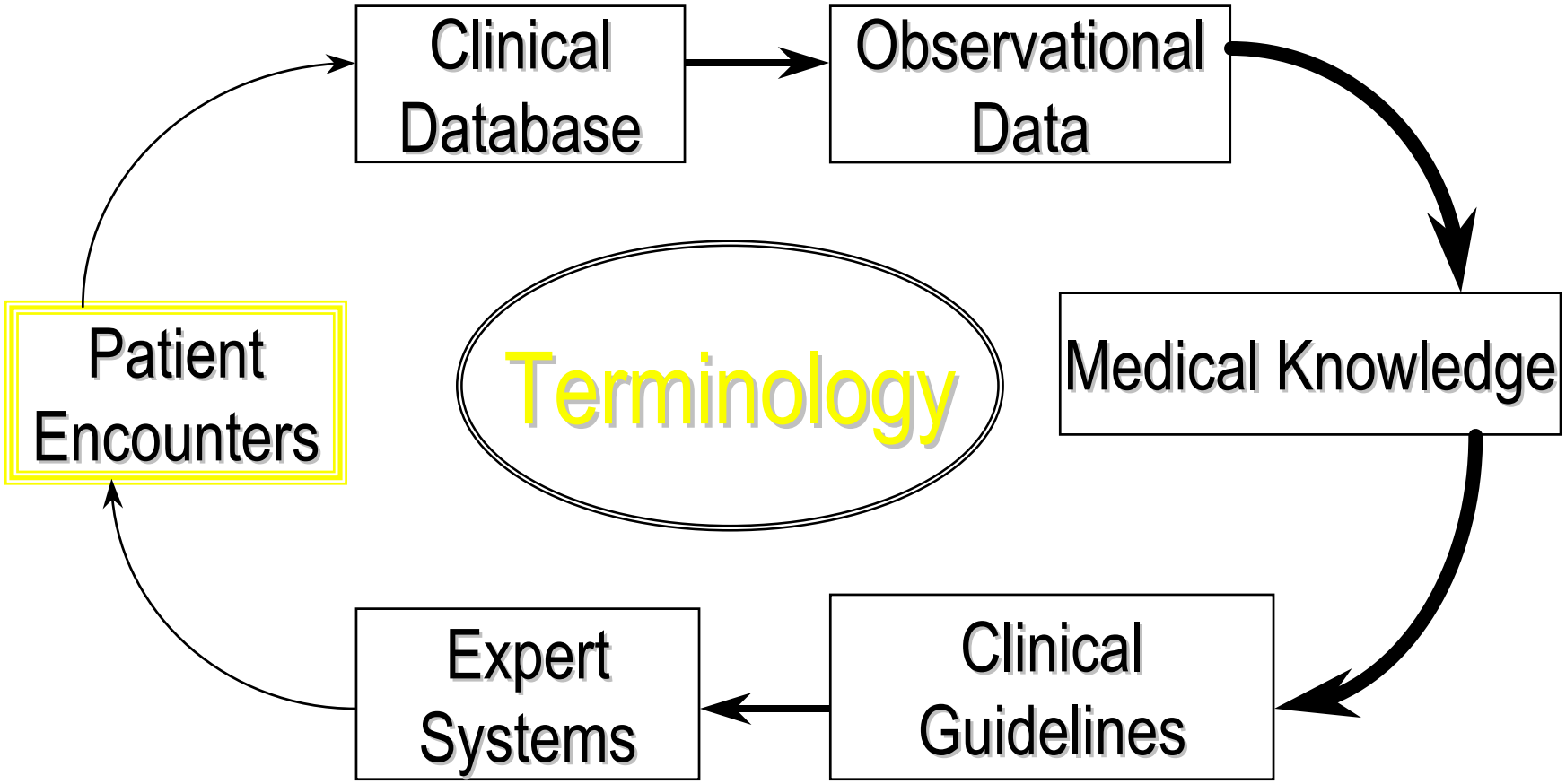
- Incomplete or inconsistent knowledge of classification rules
- Clinician resource time is scarce – billing record is often perceived as a part of the clinical record
- Indexing is resource intensive
- Indexing process depends on what is known at the time
- Researcher has to have intimate understanding of all parts of the process

Reference Terminology

Responsible for:
1) Content of medical record



Heritage of Continuous Improvement



MAYO CLINIC

Reference Terminology

- Must represent fine level of clinical detail
- Coverage must be broad enough to span an entire discipline
- Must be well defined
- Must be compositional in nature
  - Post-coordination rules
  - Rules for determine compositional equivalence

MAYO CLINIC

Characteristics of a Reference Terminology

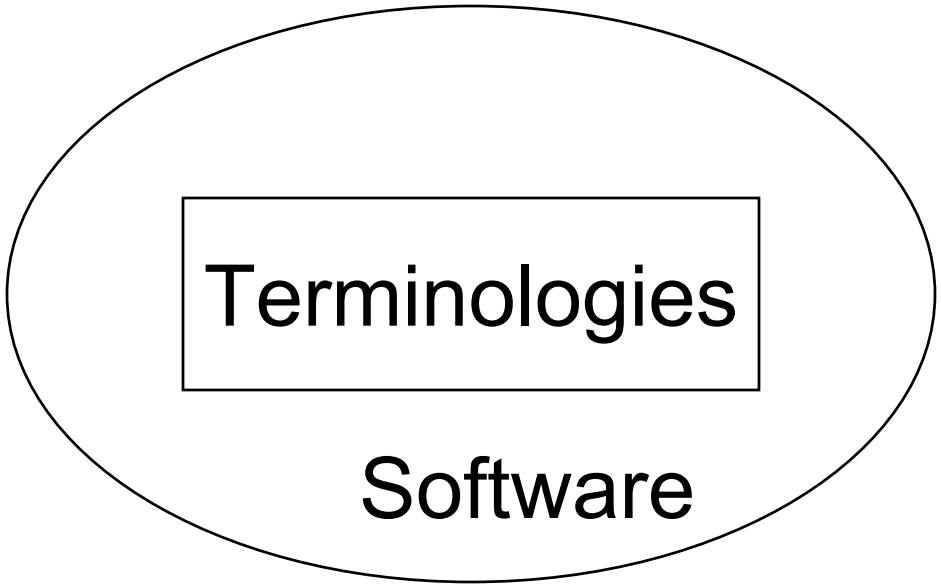
- Requires computers and software to be effectively managed (3rd generation system)
- Terminology = content + software

Requirements for Specifying Software Behavior

1. **Information model –** What are the entities that are manipulated and how are they related?
2. **Requirements model –** What questions will the software need answer?
3. **Behavioral model –** How do request various behaviors, what do they do and how do they respond?

MAYO CLINIC

Reference Terminology



MAYO CLINIC

Where we've been

Healthcare Data Dictionary
(Stan Huff / 3M)

- Evolved from the HELP PTXT system
- Software included:
  - Oracle DB running on Unix
  - Tuxedo transaction management system
  - OLE / C++ based client object system
- Reasonably successful
- Proprietary – content and algorithms didn't port

Lexicon Query Services (LQS)

- Developed under auspices of the Object Management Group (OMG)
- Read-only – no authoring
- Included:
  - Information Model of Terminology
  - Behavioral Model
  - Implementation Specification targeted for the Common Object Request Broker Architecture (CORBA)
    - Syntax specification in IDL
    - Architecture required significant changes in object/attribute layout

MAYO CLINIC

LQS Specification

- Schema for globally unique identifiers
- Validate concept code
- Lookup Concept codes by word(s) / string / pattern
- Lookup concept text for a given context / language / lexical type
- Lookup definitions / comments / instructions
- List concepts that have a specified relationship with a supplied concept
- Determine whether two concepts are related
- Reduce a composite expression to a canonical form
- Compare two composite expressions
- ...

LQS

- Published in 1998
- Not widely used or adopted (outside of 3M)
  - Perceived (and actual) complexity of the specification
  - Not easy to implement
    - Services were not trivial to implement
    - No reference implementation
    - CORBA was difficult and expensive to work with
- Model and functional requirements are still reasonably definitive

MAYO CLINIC

Mayo Terminology Services (MTS)
(Chris Chute / Mayo)

- Extension of Lexicon Query Services
- Purpose was to provide a complete “breadboard” of terminology components
- Added lexical/linguistic capabilities
  - Spell correction (word locator)
  - Word stemming (using LVG)
  - Word and Phrase completion
  - Plesionymy (words and phrases that could have a very similar meaning in a given context)
  - Candidate term matching

MTS 2000 Implementation

- Java based
- Used JDBC back end (Oracle / Sybase)
- Used SNOMED-RT Database model
- Multiple Implementation Architectures
  - CORBA
    - COM/DCOM bridge
    - Perl Bridge
    - Python Bridge
  - RMI
  - Straight Java Objects

MTS - Lessons Learned

Usefulness within Clinic depended on adoption by internal and external software providers

- Y2K came first
- Waited for HIPAA regulations to solidify
- Needed today's external incentives
  - HIPAA
  - Terrorism Surveillance
  - Bioinformatics

MAYO CLINIC

MTS Lessons Learned

Usefulness within Mayo Clinic depended on adoption by internal and external software providers

- Specification needs to be
  - Standards based
  - (Relatively) easy to implement

MAYO CLINIC

Where we are today

MAYO CLINIC

Where we are today

- Open Terminology Services
- HL7 CTS Specification
- HL7 Terminology Tools
- LDAP Back End
- OTS Using LDAP and Lucene
- NLP processing of medical records and term source

MAYO CLINIC

Open Terminology Services

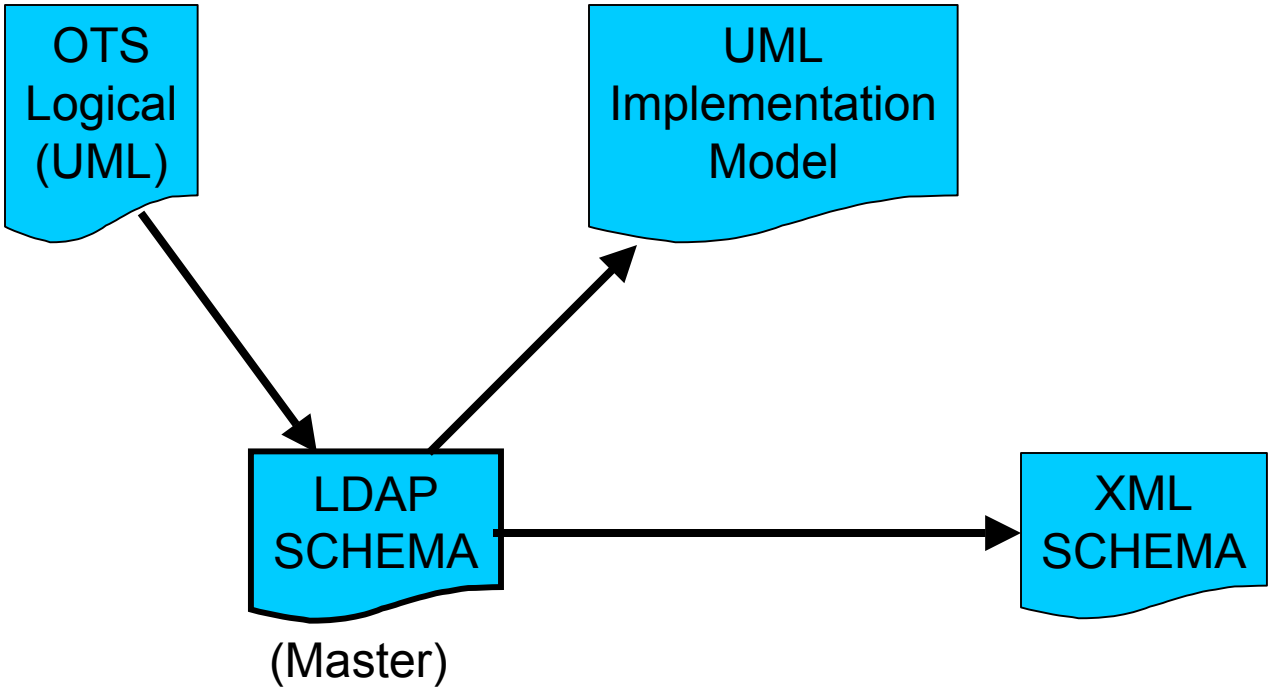
- Refactoring of MTS
- Standards Based:
  - CTS through HL7
  - “Open Source” approach
    - Java Reference implementation
    - LDAP back end
    - ... serious clash w/ Mayo culture

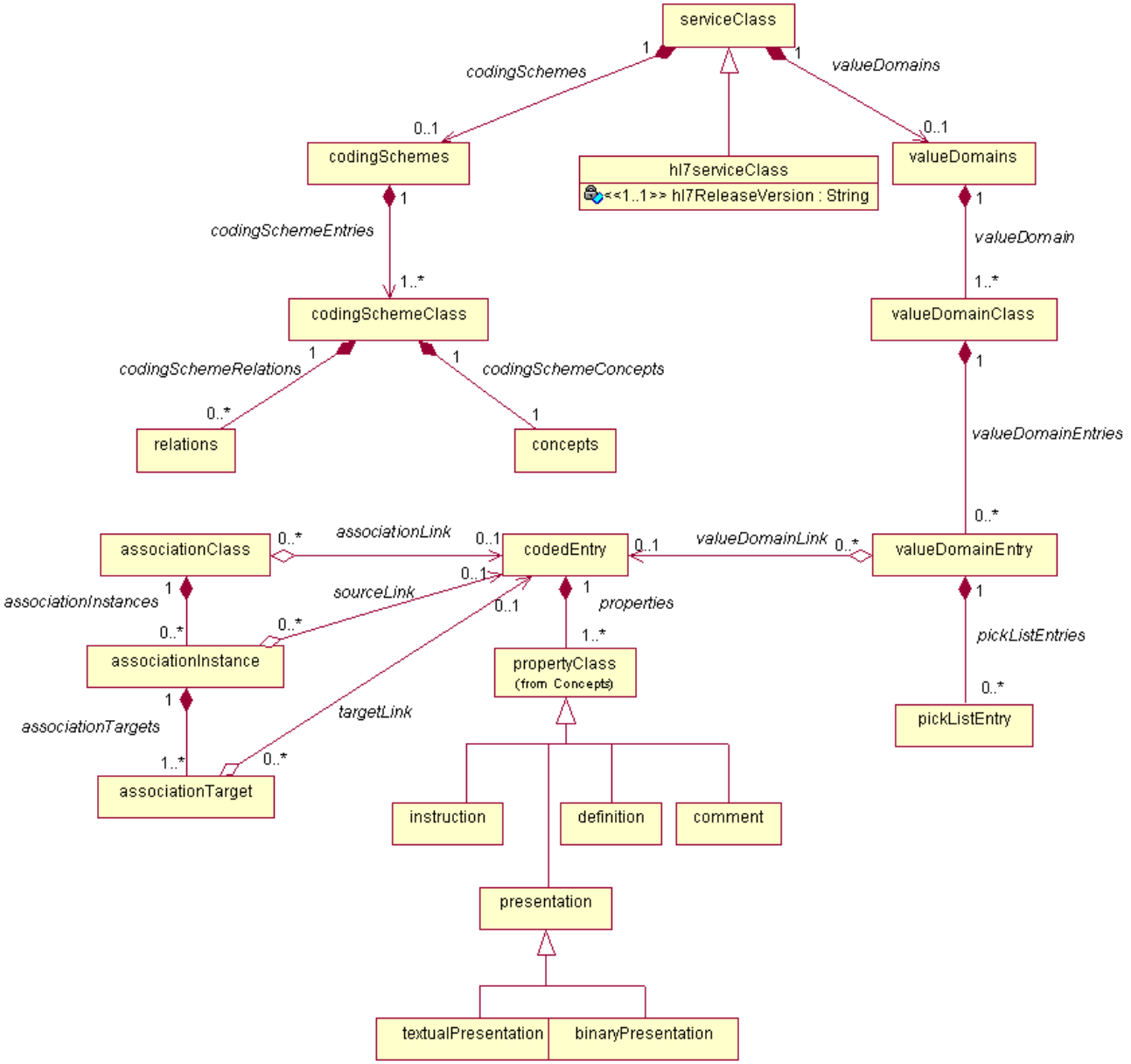
MAYO CLINIC

OTS Components

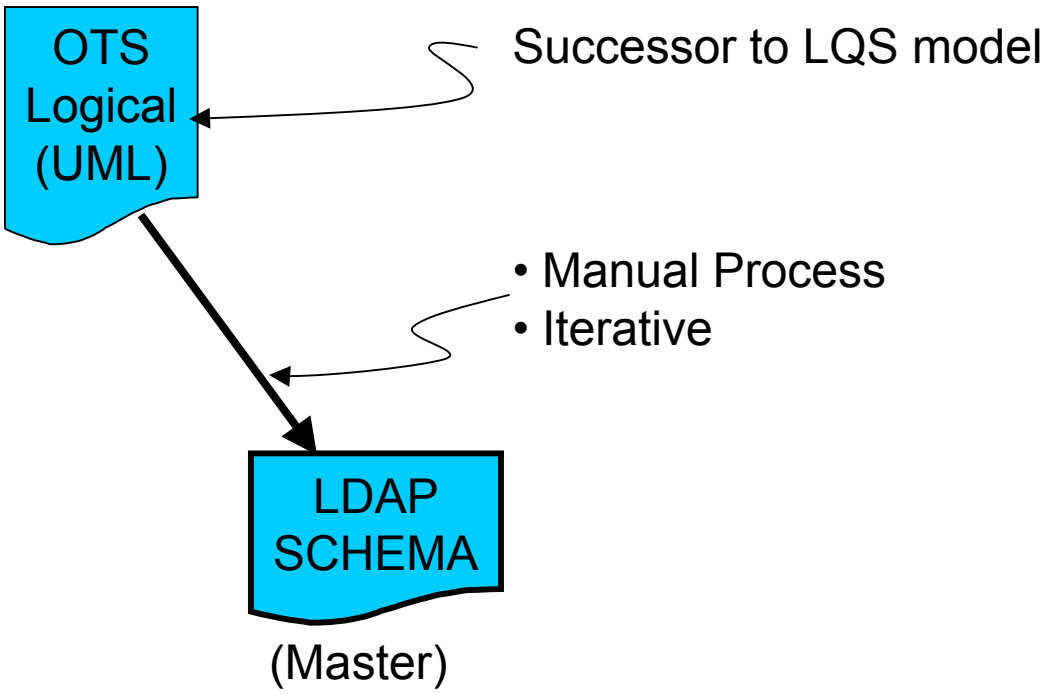
- Model
  - Abstract ('logical') model of information content
  - Implementation models – one per technology
- Content
  - Terminology content deployed in various technologies
- Software
  - Browsing and Implementation Tools
  - Distribution and deployment tools
  - Editing and revision tools

OTS Model

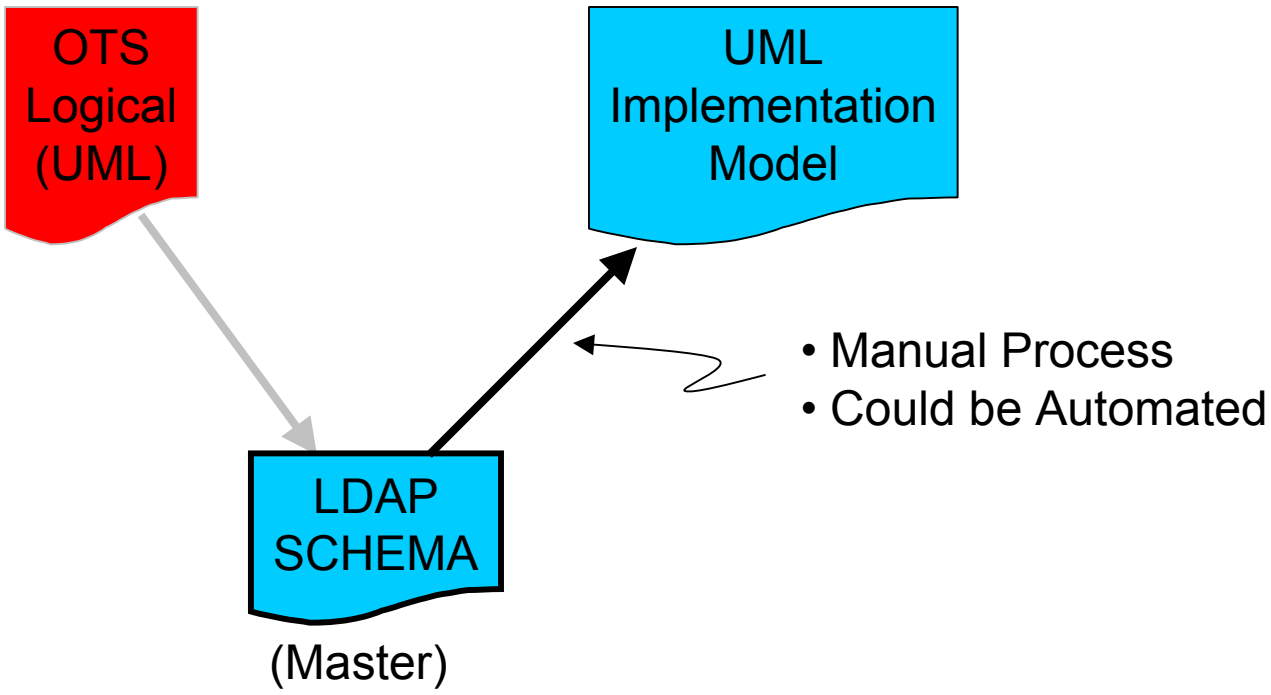




OTS Model
Logical Model & Schema

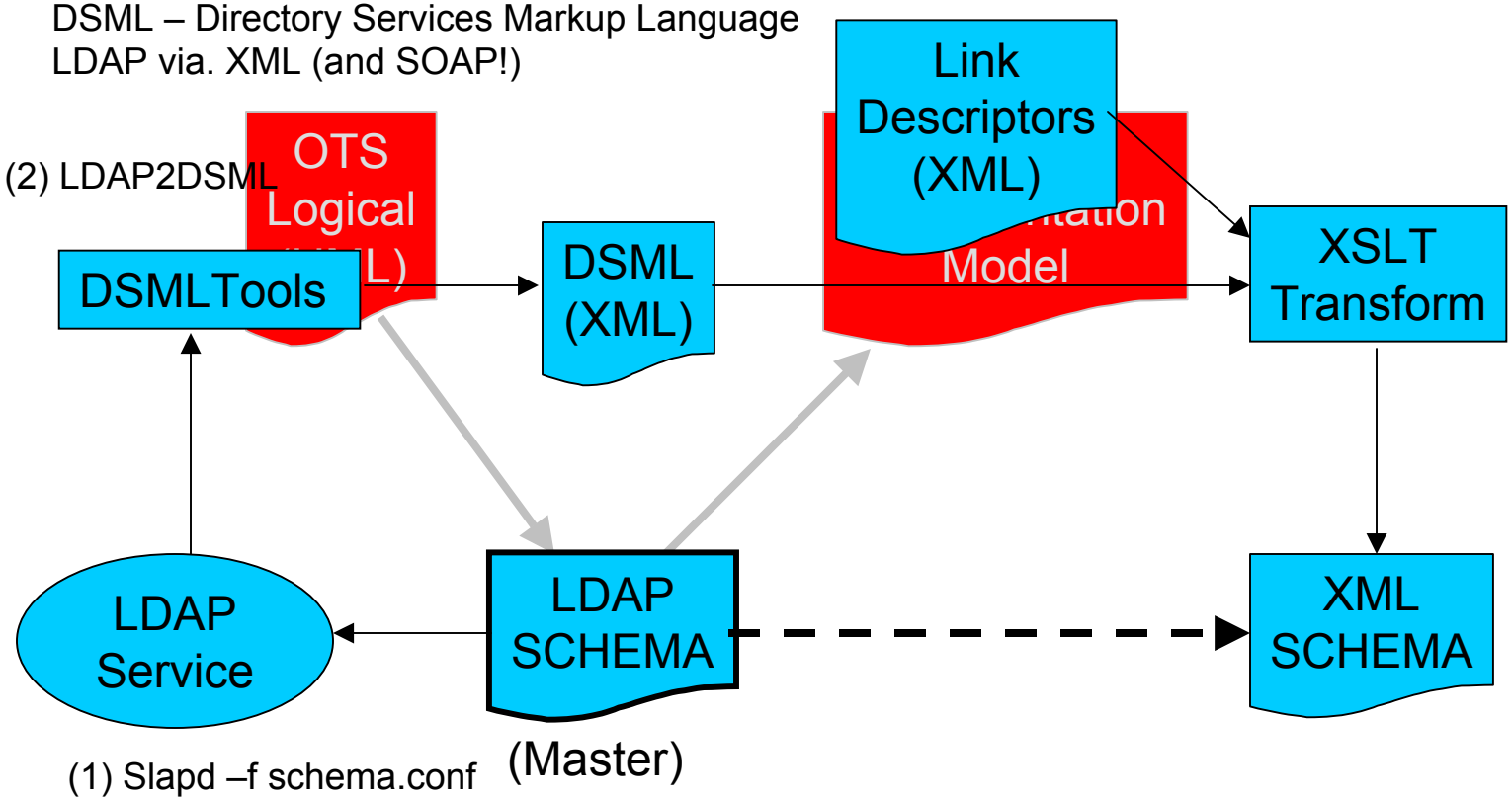


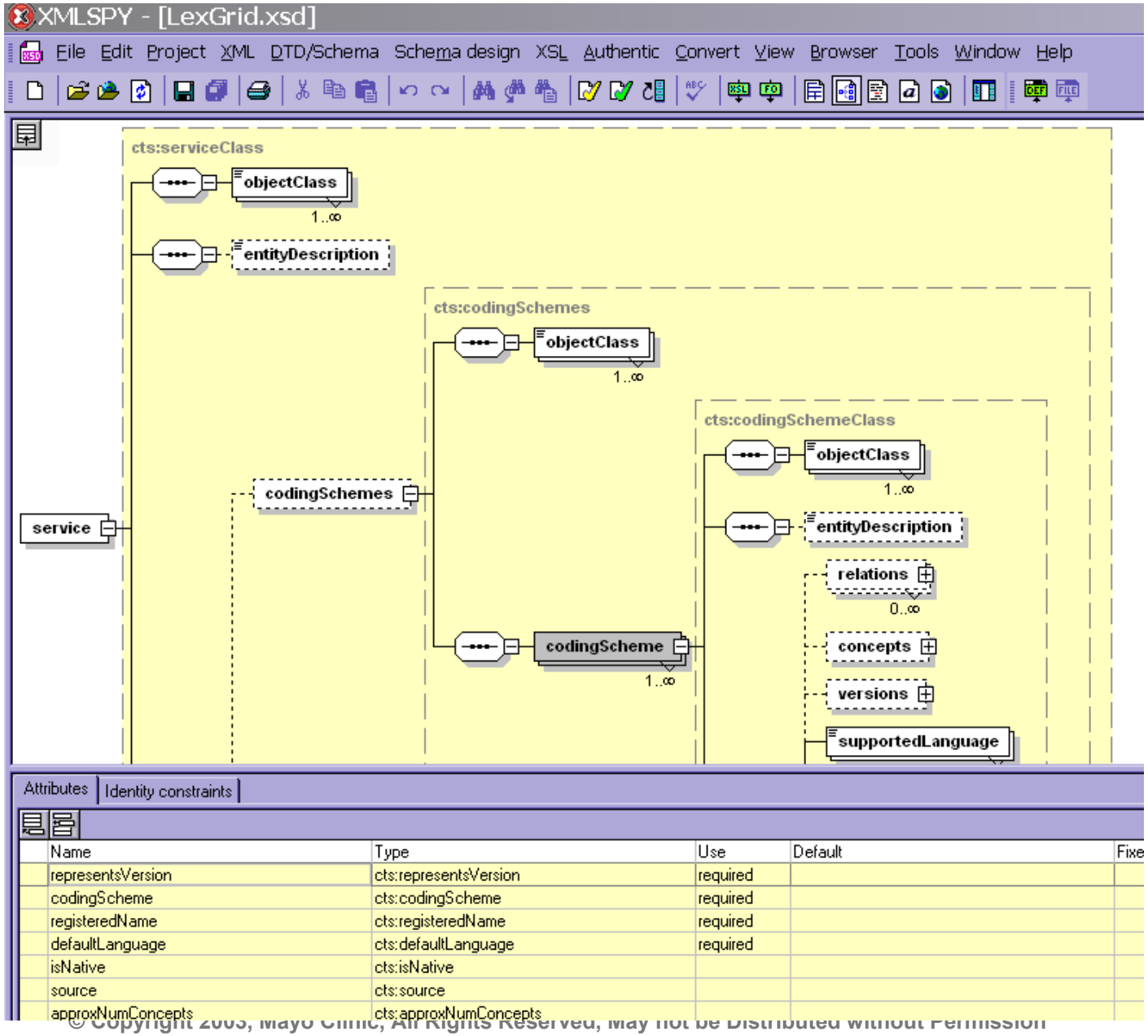
OTS Model
Implementation Model



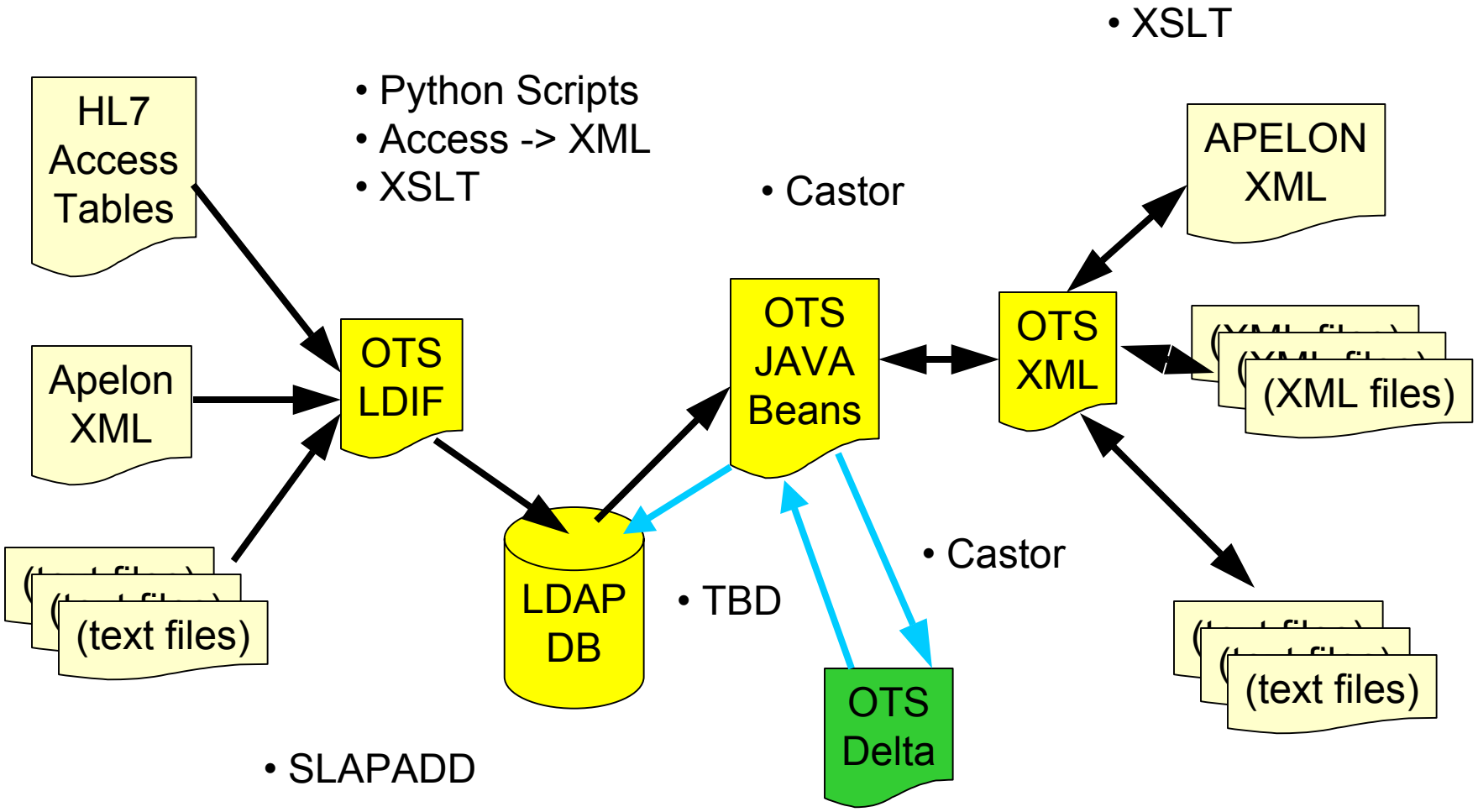
OTS Model XML Schema

DSML – Directory Services Markup Language
LDAP via. XML (and SOAP!)

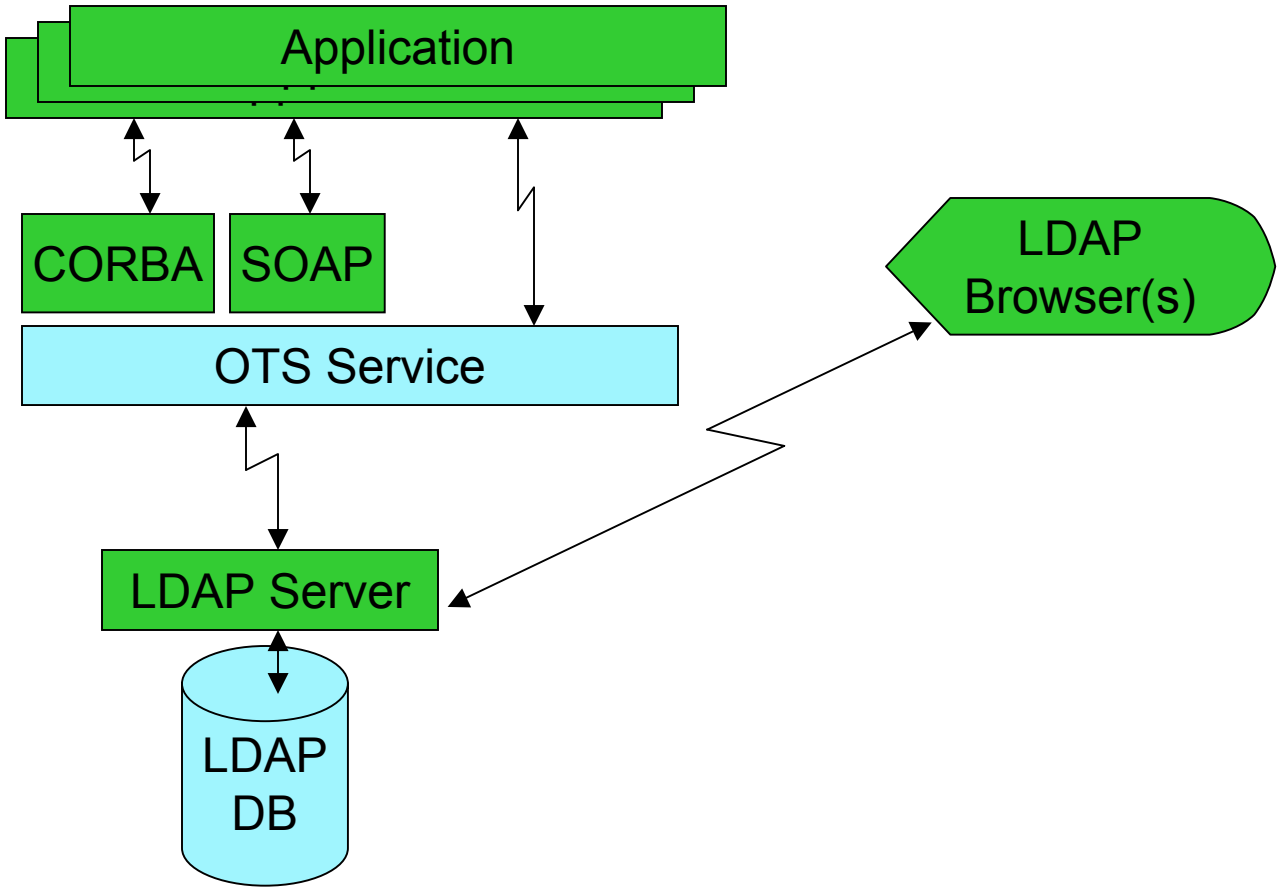




OTS Content



OTS Software
Browsing and Implementation Tools



LDAP Back end

- Lightweight Directory Access Protocol
- Used for publishing read-mostly, high-availability directories of “things”:
  - People
  - Resources
  - Organizations
  - Java Services
  - ...

MAYO CLINIC

LDAP Characteristics

- Hierarchical directories of information
- Focus is read-mostly information
- High availability, high reliability
- Supports data replication
- Reasonable security model
- Supports distributed hierarchies (federation)
- Both open source and commercial tools are widely available

MAYO CLINIC

Why LDAP?

- Widespread availability
- The hope is that vendors will find it easier to load their own content into a generic data model than it would be to write a service implementation themselves.

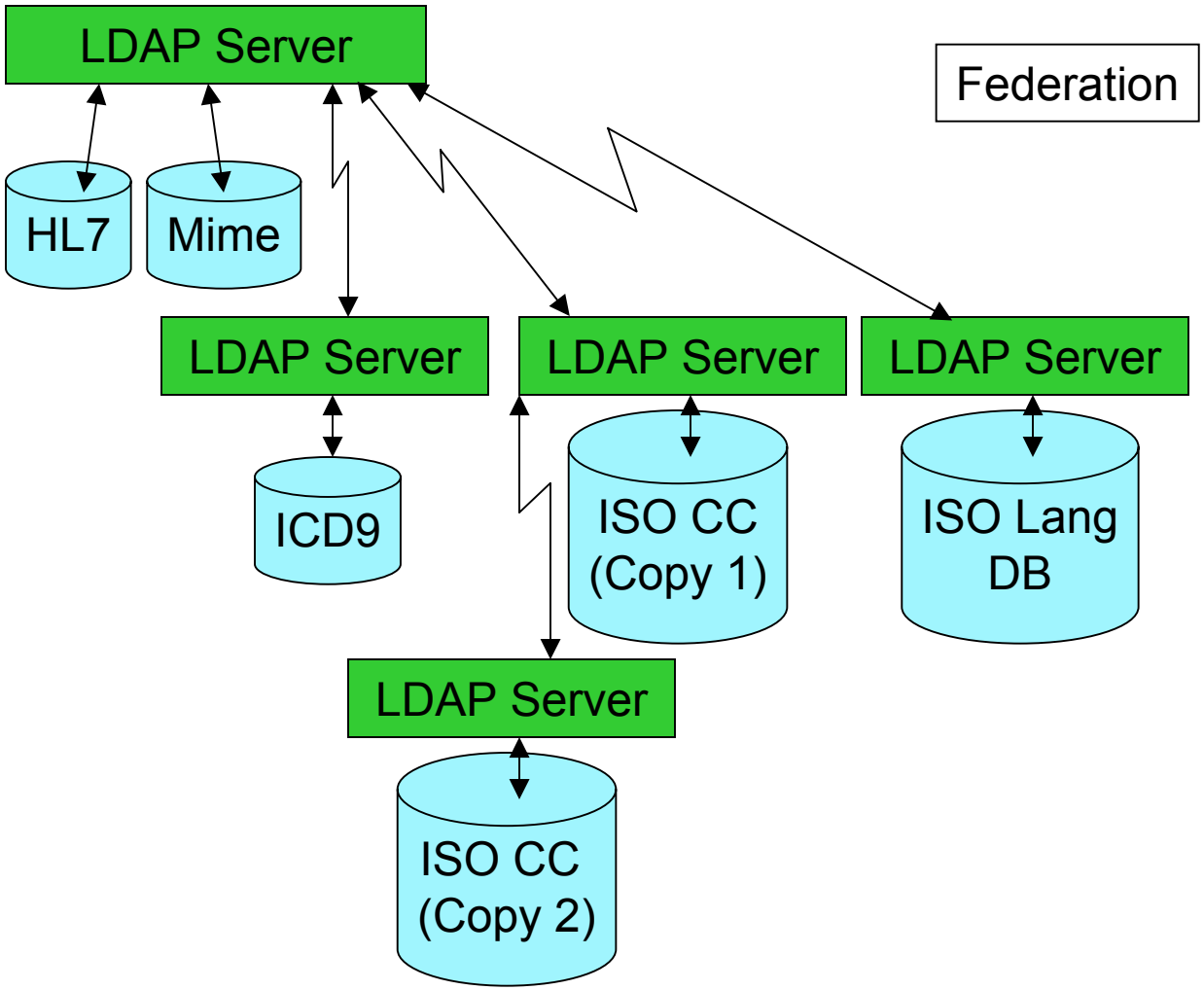
MAYO CLINIC

LDAP Back End

- Currently publishing HL7 & related terminologies
- Software & Demo can be found at

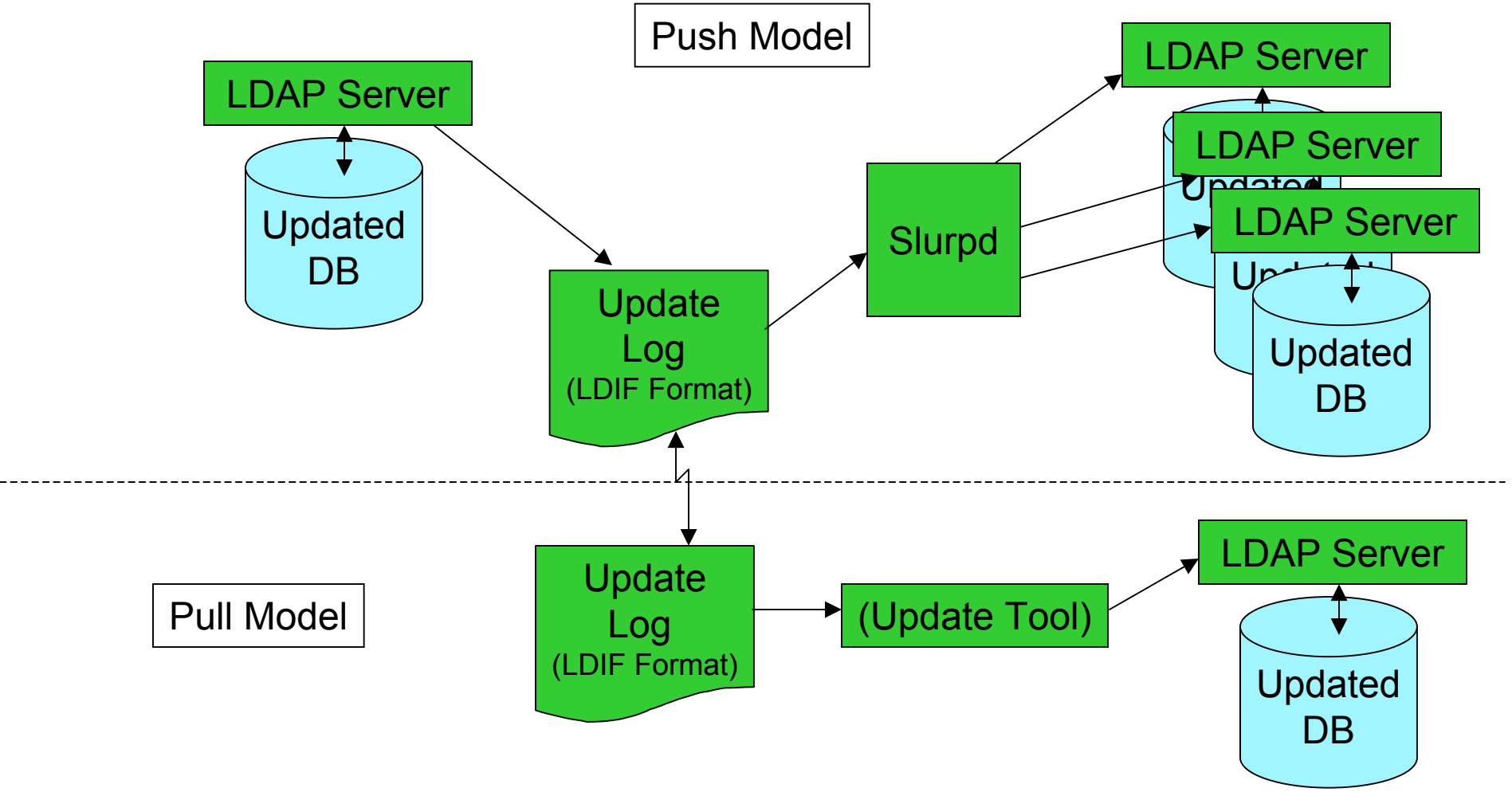
  http://www.terminologyservices.org

OTS Software Distribution and Deployment

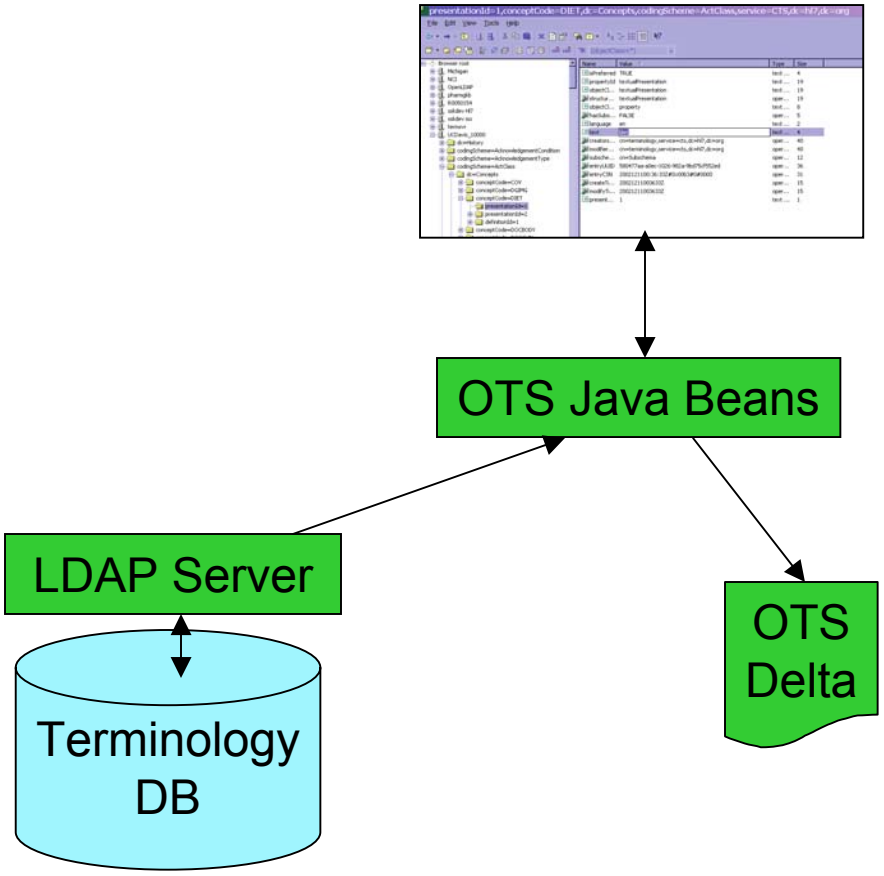


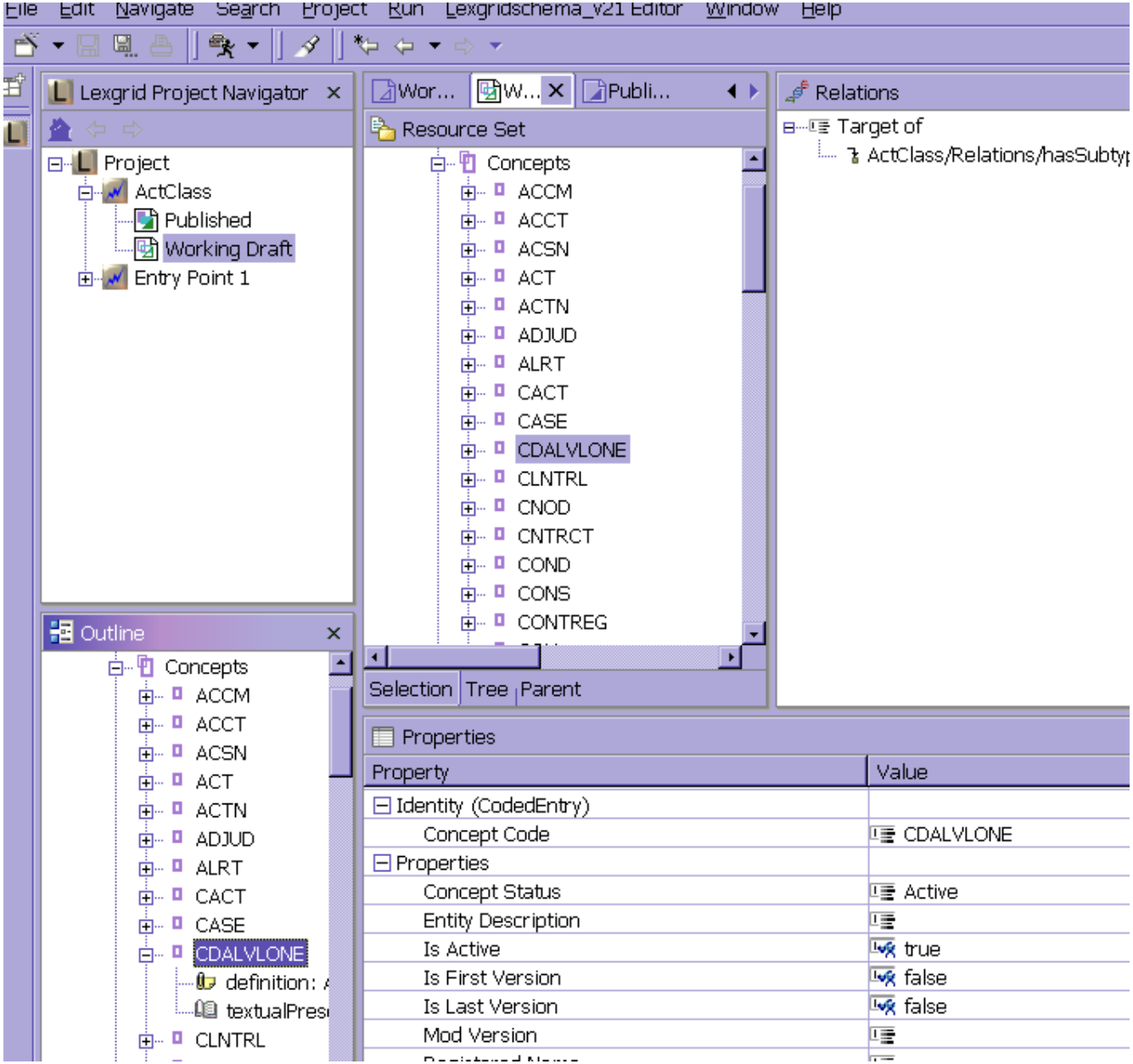
MAYO CLINIC

OTS Software Distribution and Deployment

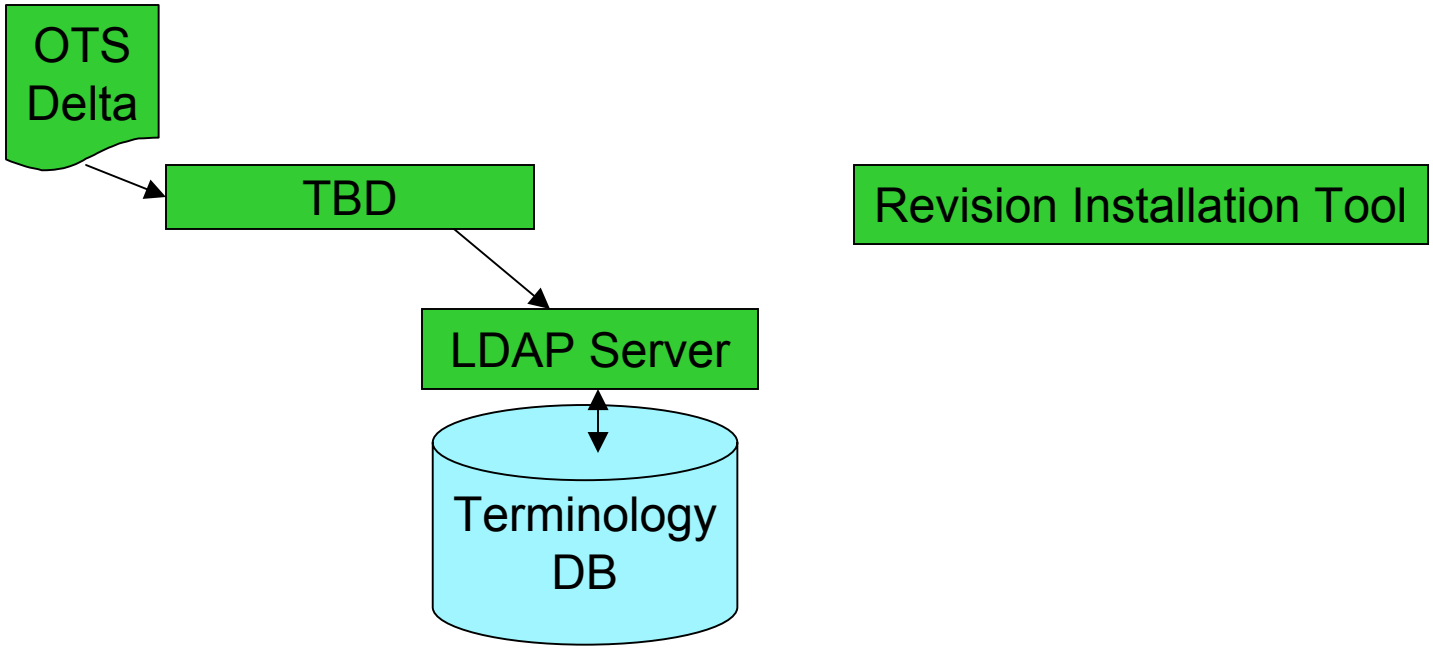


OTS Software Editing





OTS Software Revision



MAYO CLINIC

Open Terminology Services MTS Extensions

- Creating new indexing strategy w/ distributed LDAP back end
- Lucene based
- Enhancing thesaurus with various semantic distance algorithms
- Alpha should be available shortly

MAYO CLINIC

HL7 CTS Specification

- Specification divided into two parts:
  - Messaging layer – speaks HL7 messages, data types & process
  - Vocabulary layer – speaks code systems / terminology
- Still under revision
- OMG IDL being used for syntax portion
- Having to juggle needed functionality and perceived simplicity

MAYO CLINIC

HL7 CTS Specification "Reference" Implementation

- Native Java or SOAP based
- Messaging API uses JDBC back end
- Vocabulary API uses LDAP back end
- Demonstration code (0.8) and source available on web

  http://www.terminologyservices.org

MAYO CLINIC

HL7 Terminology Tools

- Supporting HL7 Vocabulary Maintenance
- XML-Based Submission Format
- Processor and update tool
- Still need to reintroduce historical part
- Editing tools pending:
  - Apelon
  - Health Language Inc?
  - Internal editor under development
  - Protégé?

MAYO CLINIC

Open Terminology Services Content

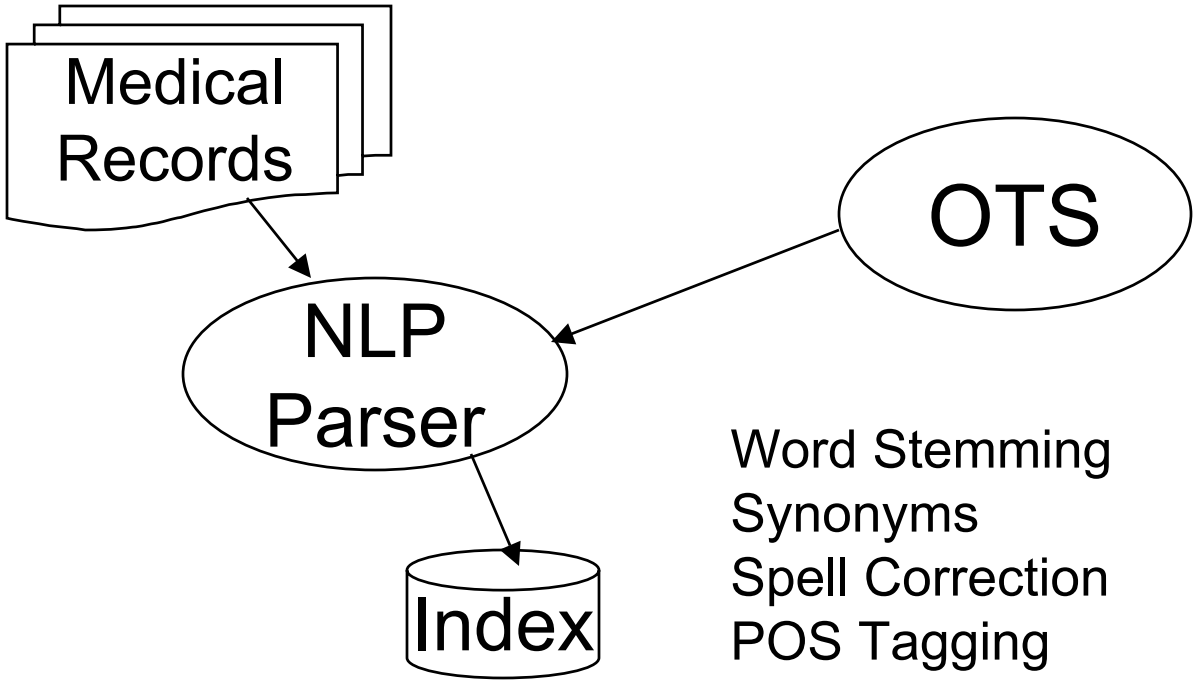
- Once implemented how do we:
  - Import and export code systems from the service?
  - Cross-reference content?
  - Post and distribute updates?
  - Edit and revise content?

MAYO CLINIC

Indexing and Cross Referencing

- Graphic of OTS & Lucene index
  - Mention of thesaurus and semantic distance stuff
  - Mention of spelling issues

NLP Based Medical Record Indexing



Word Stemming
Synonyms
Spell Correction
POS Tagging
Semantic Distance Metrics

MAYO CLINIC

Where we are going

MAYO CLINIC

Vision

- The Lexical Grid
- Blurring the Terminology / Information Model Boundary
- Terminology on the front end

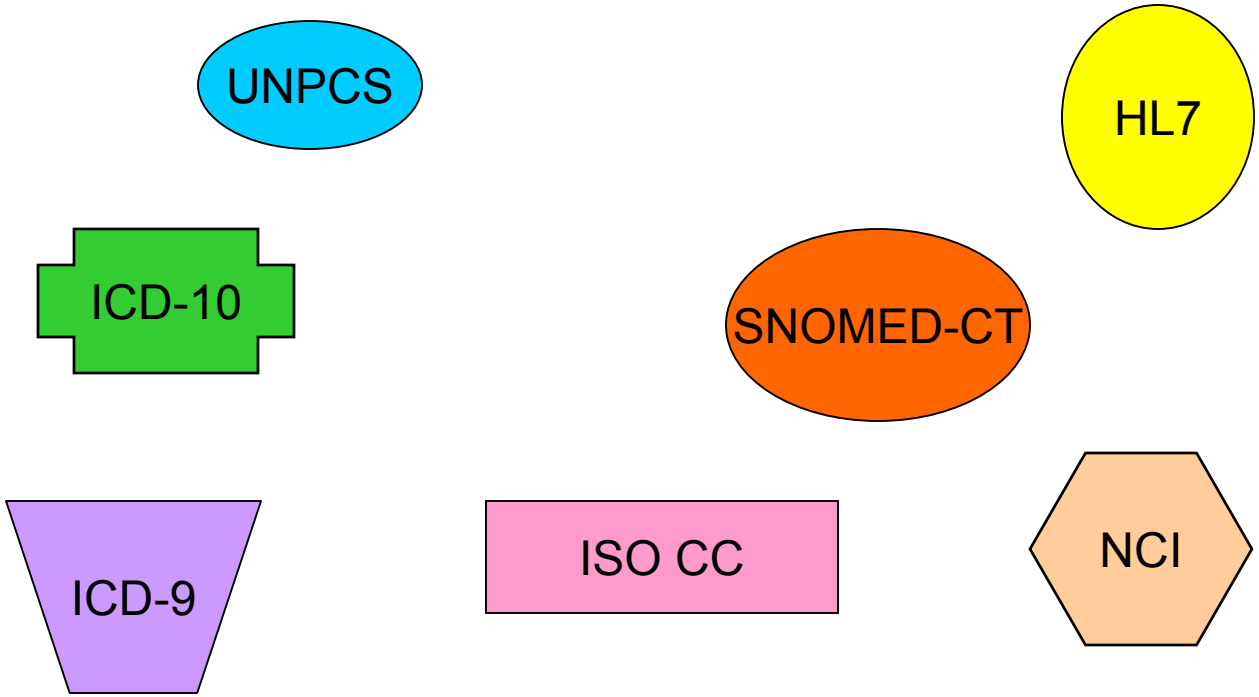
MAYO CLINIC

The Lexical Grid

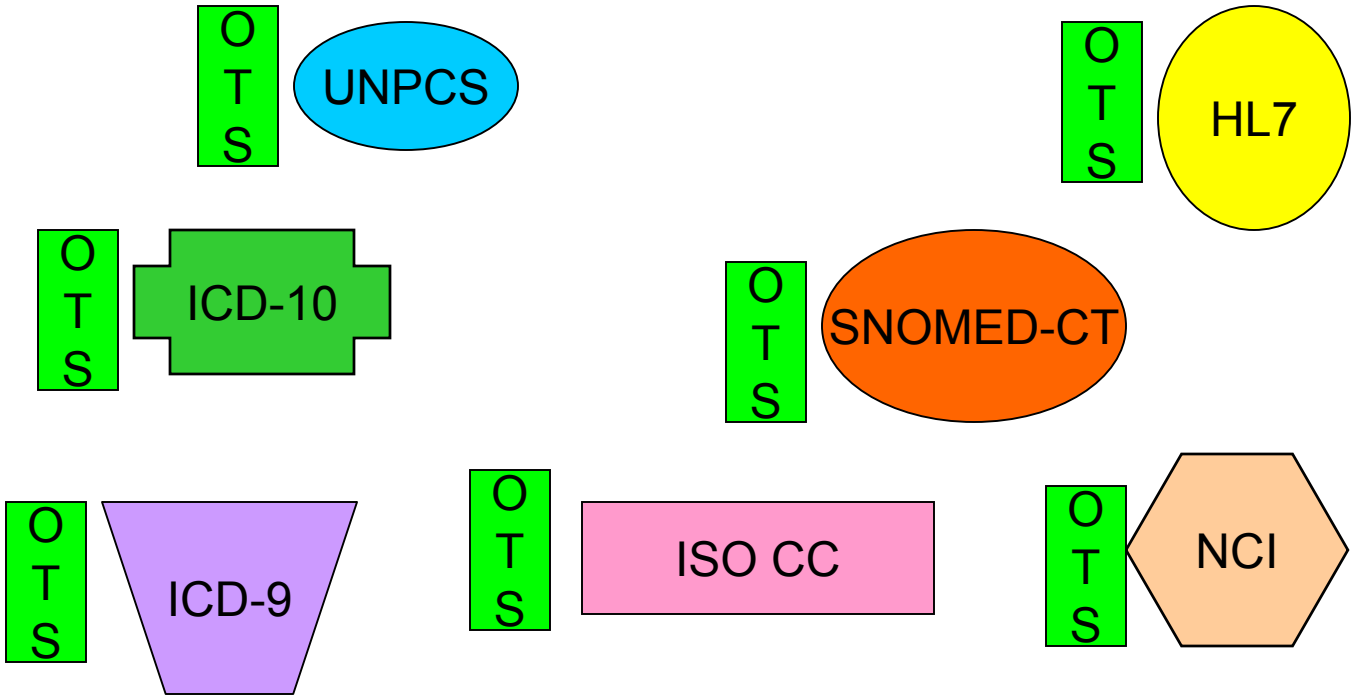
A heterogeneous, distributed collection of terminologies...

- ...linked by a common API
- ...coupled to shared indices
- ...capable of being extended, enhanced an annotated in a loosely-coupled, distributed fashion

A Heterogeneous, Distributed Collection of Terminologies

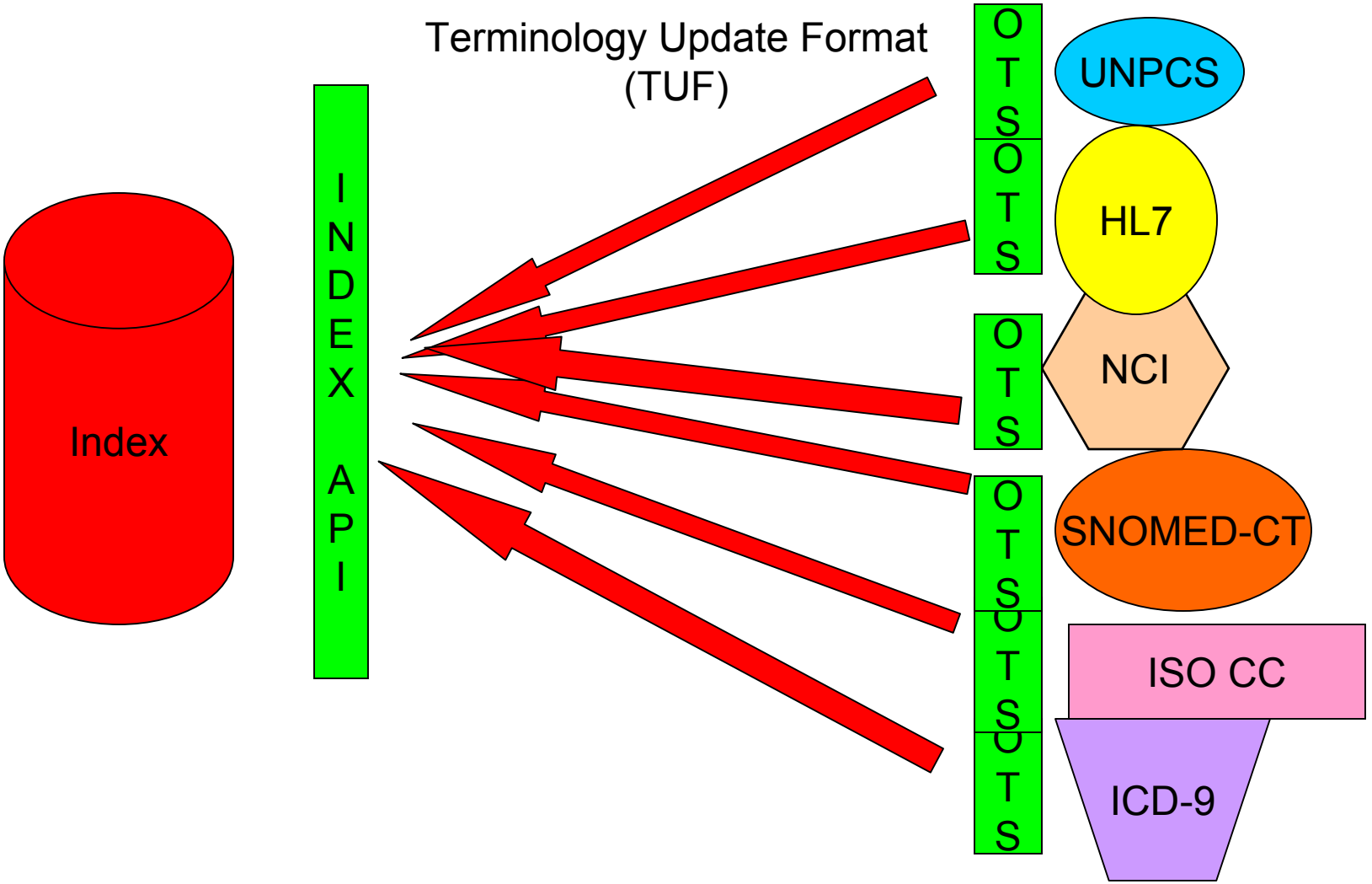


Linked with a Common API

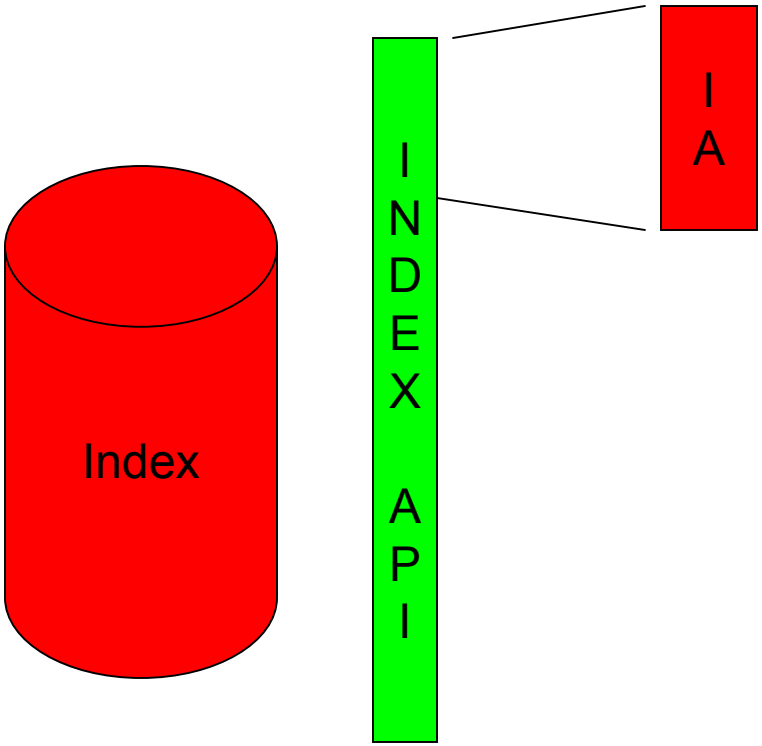


MAYO CLINIC

Coupled to Shared Indices



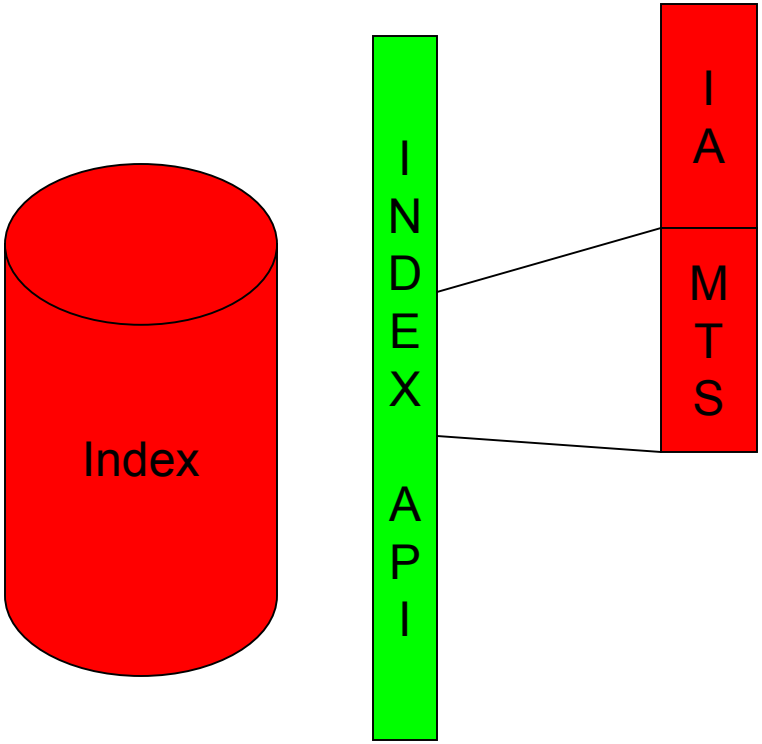
Index API
Generating an Index



Index This (Id, {(type, text)}
Update Index(Id, {type,t ext)}
Unindex (Id)
Unindex(part ID)

ID – URN
part ID - entire terminology
Need URN -> Service Map

Index API
Query Interface

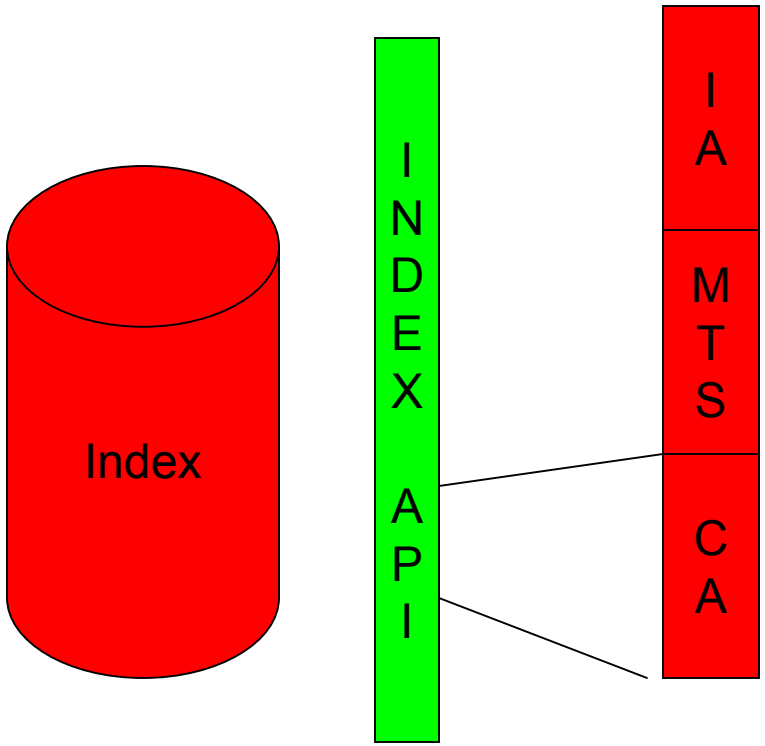


MTS ++
List ID's (code system/concept)
Matching(phrase, semantic type, etc)

Query -> Id list (csid + code)

MAYO CLINIC

Index API
Consolidated API

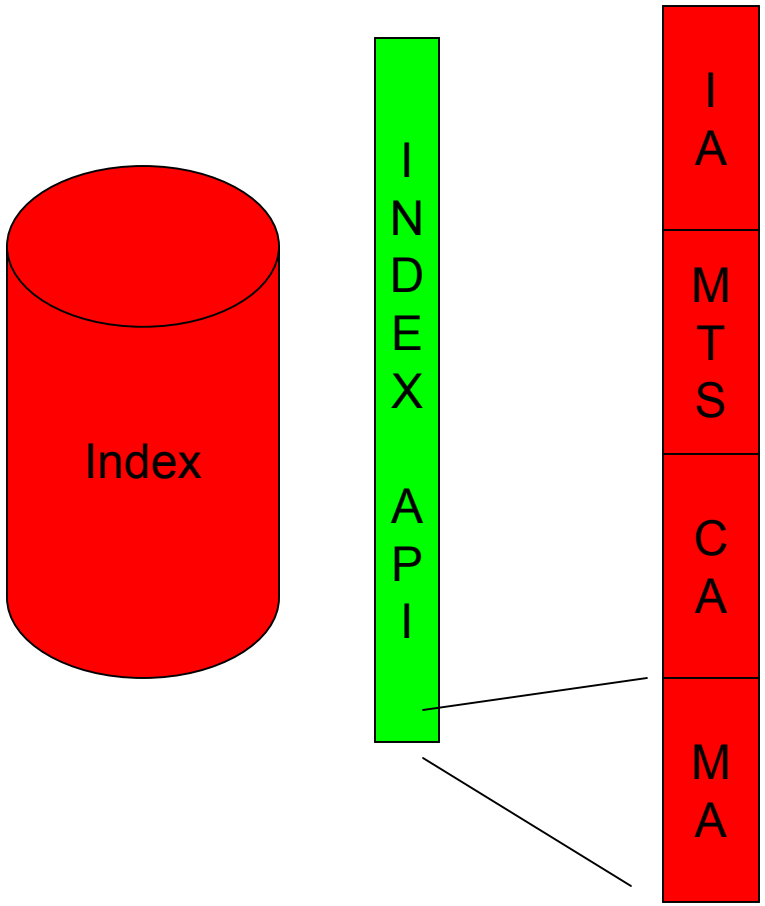


Presents Terminological Space as a Single entity

Consolidated OTS Query – user has no need to query individual vocabularies

MAYO CLINIC

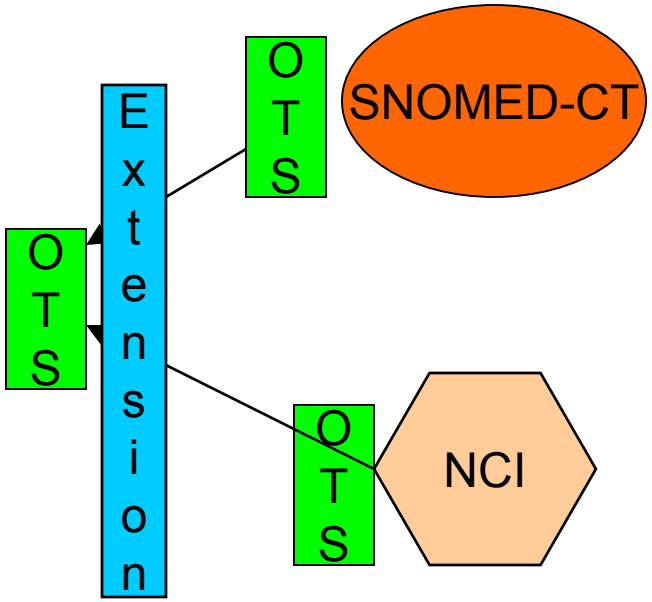
Index API
Maintenance API



ManageThesaurus
Spelling Dictionaries
Other indexing packages

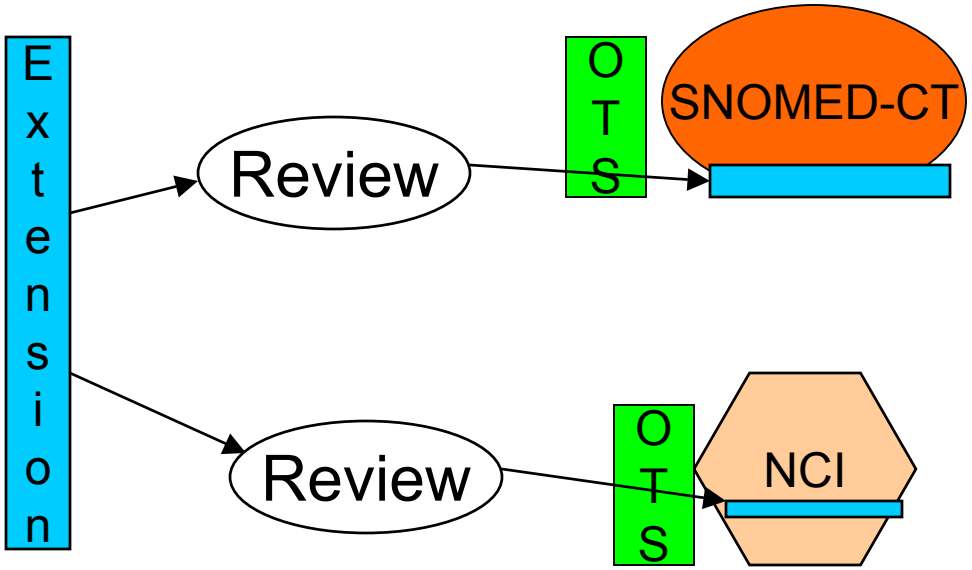
MAYO CLINIC

Extended, enhanced an annotated in a loosely-coupled, distributed fashion



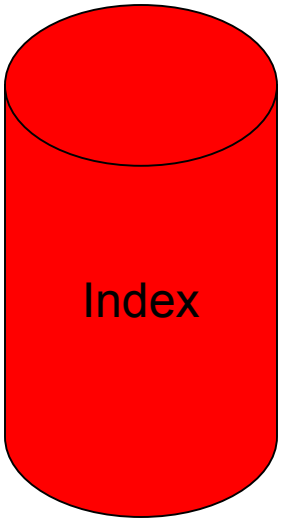
MAYO CLINIC

Extended, enhanced an annotated in a loosely-coupled, distributed fashion



MAYO CLINIC

Index

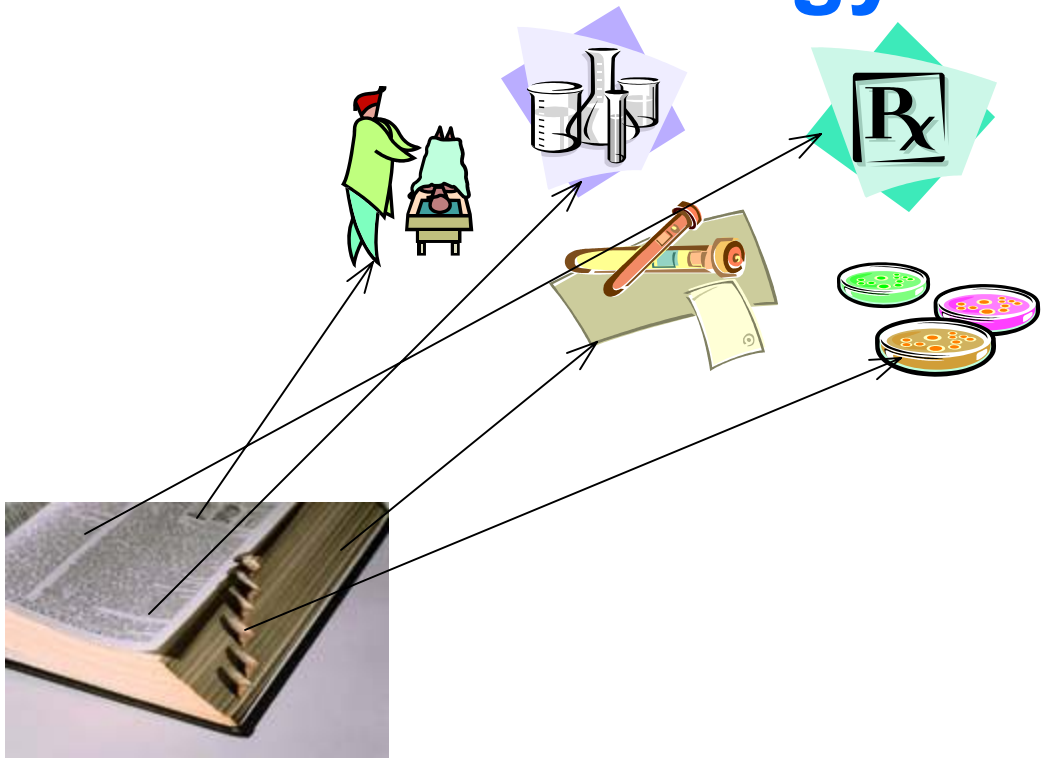


- Misspelling tolerant
- Plesionyms
- Morphological Roots
- Phrase Library
- POS Aware
- Co-occurence info
- N-grams
- Etc...

Blurring the Terminology / Information Model

MAYO CLINIC

Terminology



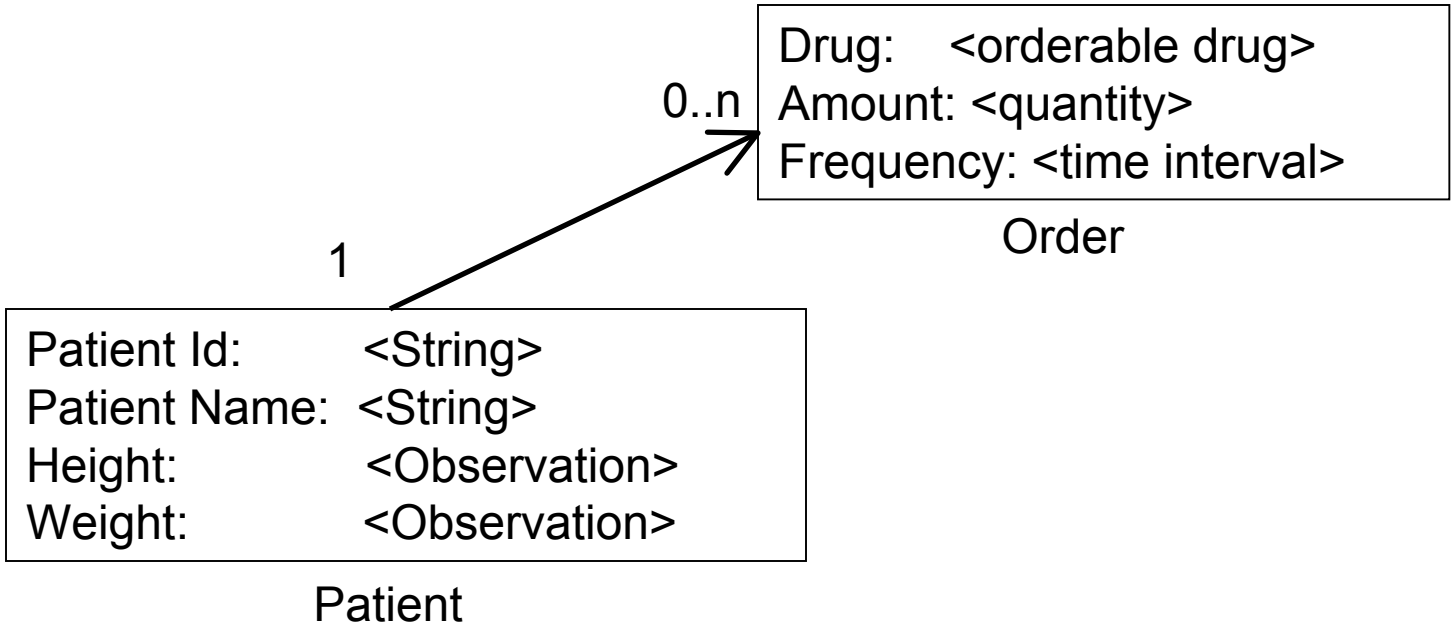
MAYO CLINIC

Information Model

- Selects the subset of the 'real world' to be discussed in a given context
- Utilizes elements in the terminology
- Tacit or explicit agreements on what is to be:
  - Ignored
  - Refined
  - Expanded and augmented
- Extends the terminology model with non-definitional characteristics

MAYO CLINIC

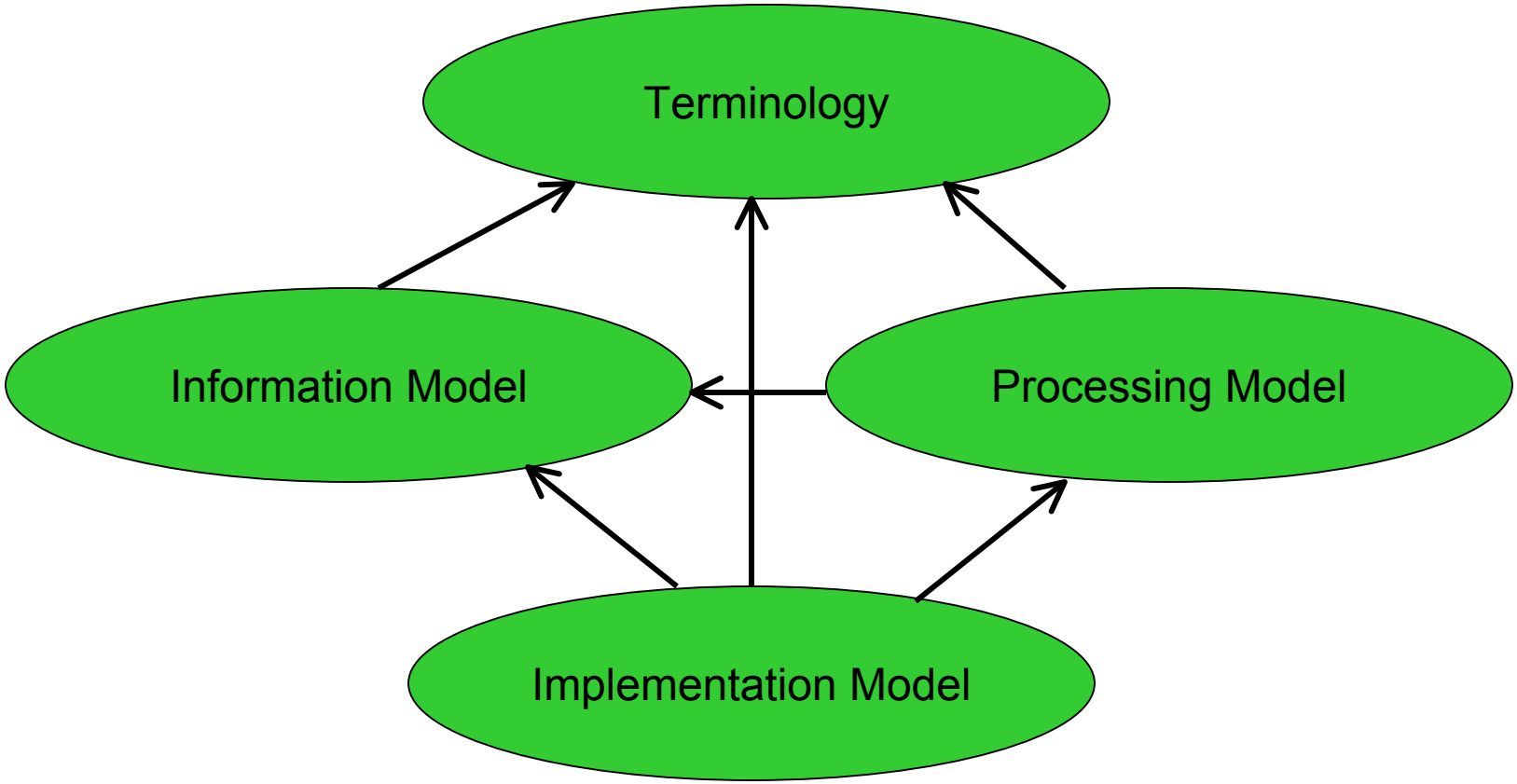
Information Model



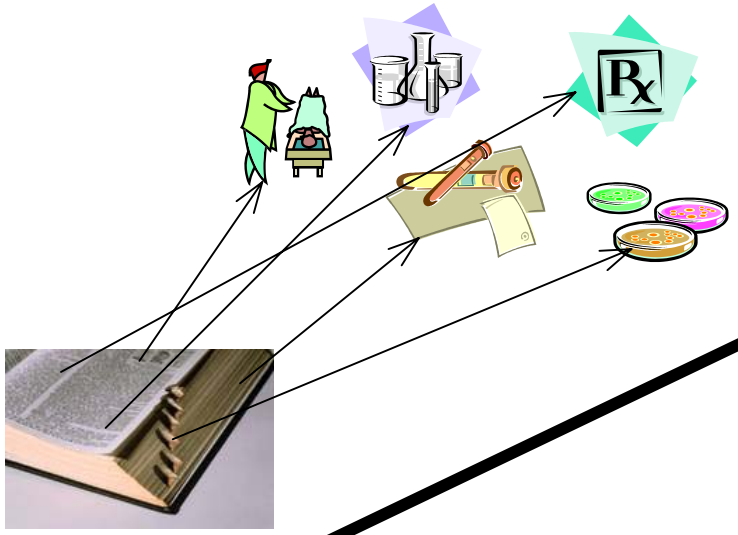
Pharmacy Orders

MAYO CLINIC

Dependencies
(Borrowing heavily from RM-ODP)



How Do We Link...



Drug: <orderable drug>
Amount: <quantity>
Frequency: <time interval>

Order

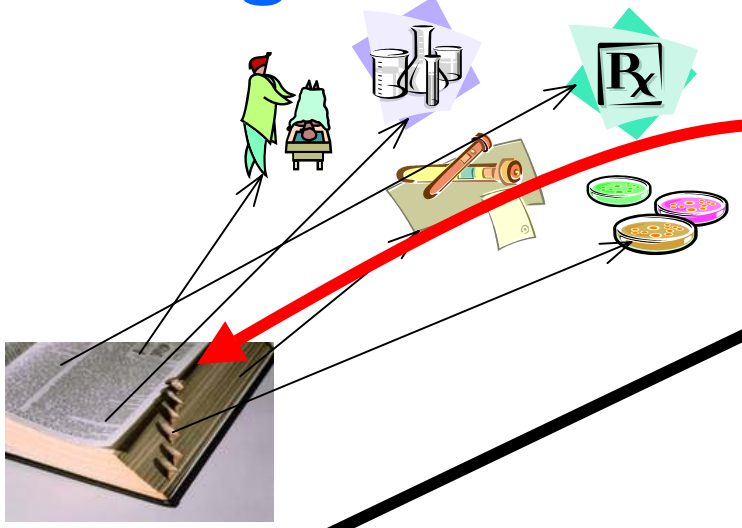
Patient Id: <String>
Patient Name: <String>
Height: <Observation>
Weight: <Observation>

Patient

Pharmacy Orders

MAYO CLINIC

Linking at the Attribute Level



Drug: <orderable drug>
Amount: <quantity>
Frequency: <time interval>

Order

Patient Id: <String>
Patient Name: <String>
Height: <Observation>
Weight: <Observation>

Patient

Pharmacy Orders

MAYO CLINIC

Database Granularity

The same information can be carried in widely varying structures:

A code in a table

PT#	observation
1110112	**Heart murmur**

Tag/Value Pairs

PT#	Tag	Value
1110112	**Observation**	**Murmur**
1110112	**Location**	**Heart**

Free text

PT#	observation
1110112	**"Findings indicate a pronounced apical murmur throughout the entire systolic phase…"**

Column Headings

PT#	Heart Murmur
1110112	**TRUE**

Table Names

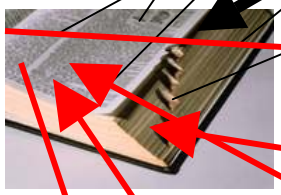
PT#
1110112

Table 17: Patients with Heart Murmur

Database Names

MAYO CLINIC

Linking at All Levels



Drug: <orderable drug>
Amount: <quantity>
Frequency: <time interval>

Order

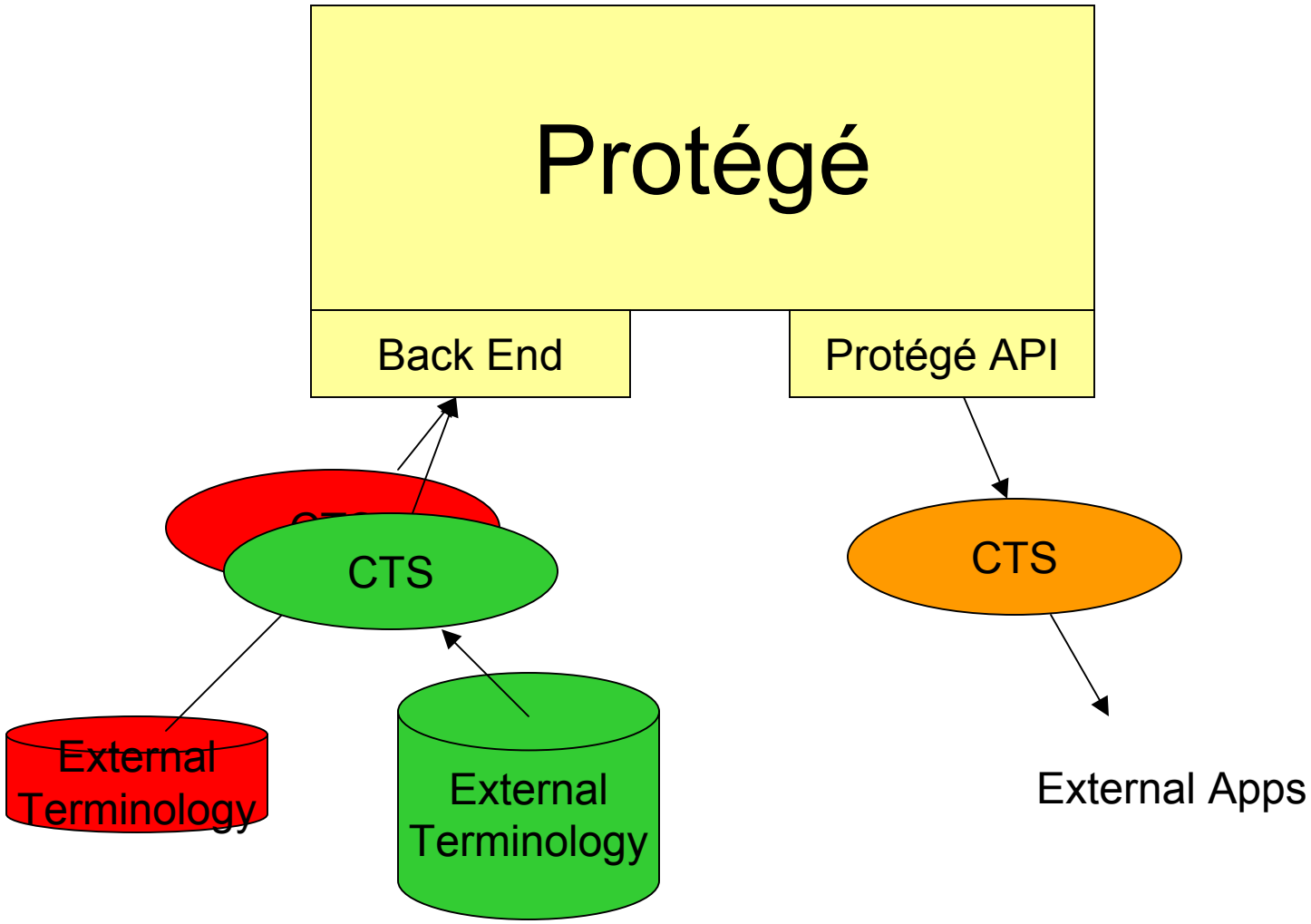
Patient Id: <String>
Patient Name: <String>
Height: <Observation>
Weight: <Observation>

Patient

Pharmacy Orders

MAYO CLINIC

Proposed Approach



MAYO CLINIC

Protégé

- Promoted as an ‘ontology’ editor
- Frame based – no integrated classifiers
- Open source / open architecture
  - > 50 User written add ons
    - Graphical Browsers
    - Reasoning Engines
    - Back Ends
    - Different Input and Browsing forms
  - Currently being integrated w/ OWL

Protégé and CTS/OTS

1. **Protégé needs a strong terminological link**
   - **Words or sets_of_words within an ontology are not sufficient**
   - **Need to be able to reference and reuse wherever possible**
   - **External terminologies should be available in the Protégé metaphor**

MAYO CLINIC

Protégé and CTS/OTS

2. **Protégé authored ontologies need to be accessible to a wide variety of applications**
   - **Protégé API is one option**
   - **Exposing Protégé authored material via. terminology services is a second**

MAYO CLINIC

Terminology on the Front End

MAYO CLINIC

Terminology on the Front End

Tools, standards, and API's that allow the clinician to quickly, accurately enter information in an clear, unambiguous fashion

- **Spelling**
- **Phrase library**
- **Dictionaries**
- **Compositional tools**
- **Data driven forms**

MAYO CLINIC

Terminology on the Front End

Step 1: Standards

- Standard message formats
  - HL7 and derivatives
- Standard terminologies
  - OHT, UMLS++, Lexical Grid
  - SNOMED-CT (???)
- Standard plug-n-play tools
  - CTS, OTS, CCOW, ???

MAYO CLINIC

CTS / OTS Merger

- CTS – Common Terminology Services
  - HL7 specification under development
  - Attempt to balance simplicity and capability
  - Subset of OTS
- (Hopefully) Version 2.0 & OTS will be one in the same.

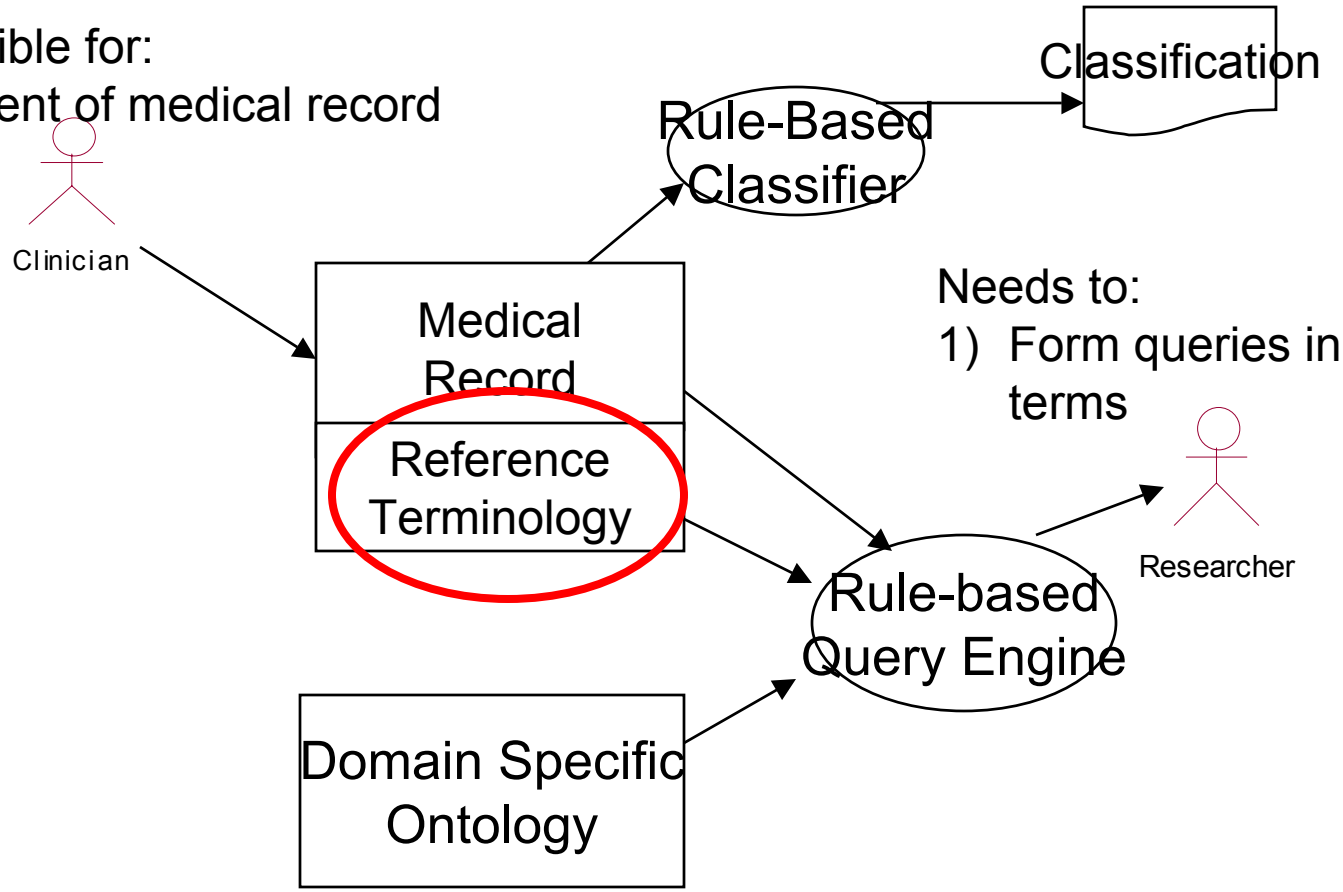
Composition and Federation

OWL and Terminology

- OWL – Web Ontology Language
  - Focus is on formal “definitions” – an entity is defined by its position in a lattice
  - Lexical aspects underspecified
    - Textual definitions and references
    - Representations / languages / contexts / linguistic forms
- Merger point may be in Protégé
- Alternative may be a transform between LexGrid model and OWL

Reference Terminology

Responsible for:
1) Content of medical record



MAYO CLINIC

Composition and Federation

- A single integrated code system is still a long way off
  - Some would argue that it will never occur
- Clinical statements will need to be made using multiple terminologies:
  - ICD-9 - ISO 639
  - LOINC - RxNorm
  - HL7 - ...
  - SNOMED-CT
  - ISO 3166

MAYO CLINIC

Composition and Federation

- The boundary between ‘information mode’ and terminological model will continue to be fluid
- How do we achieve consistent, comparable results across
  - Pre-Coordinated terms
  - Information Model
  - Terminological constructs

MAYO CLINIC

Composition and Federation

- **Part of:**

  NLM Grant 1R01LM007319-01A1

  "Development and Evaluation of Terminology Services"

Summary

- Terminology is both content and software
- Both content and software need to become widely available
  - In a variety of formats
  - For a variety of platforms
- Mayo continues to research and develop:
  - Standardized terminology service software
  - Tools for editing and distribution
  - Mechanisms for combining terminology, information models and implementation

MAYO CLINIC

Acknowledgements

NLM 1R01LM007319-01A1

"Development and Evaluation of Terminology Services"

NIST FAA 70NANB1H3049

"Standards-Based Sharable Active Guideline Environment"